Bayesian truncated beta nonlinear mixed-effects models

Citation for published version:
Paraiba, CCM, Bochkina, N & Diniz, CAR 2018, 'Bayesian truncated beta nonlinear mixed-effects models', *Journal of Applied Statistics*, vol. 45, no. 2, pp. 320-346. https://doi.org/10.1080/02664763.2016.1276891

Digital Object Identifier (DOI):
10.1080/02664763.2016.1276891

Published In:
Journal of Applied Statistics

To appear in the *Journal of Applied Statistics*
Vol. 00, No. 00, Month 20XX, 1–34

RESEARCH ARTICLE

Bayesian truncated beta nonlinear mixed-effects models

Carolina Costa Mota Paraíba[a*], Natalia Bochkina[b] and Carlos Alberto Ribeiro Diniz[a]

[a] *Departamento de Estatística, Universidade Federal de São Carlos, São Carlos-SP, Brasil*;
[b] *University of Edinburgh and Maxwell Institute, United Kingdom*

(*Received 00 Month 20XX; accepted 00 Month 20XX*)

Truncated regression models arise in many applications where it is not possible to observe values of the response variable that are above or below certain thresholds. In this paper we propose a Bayesian truncated beta nonlinear mixed-effects model, assuming that the truncated response variable follows a truncated beta distribution and that its location parameter is parametrized by a nonlinear continuous and twice differentiable function of unknown fixed parameters and covariates and by random effects. The proposed model is suitable for any response variable, y, bounded to an interval (a, b) without the need to consider a rescaled variable $y^* = (y - a)/(b - a)$ to apply the well known beta regression model and its extensions, which are primarily suitable for response variables in the unit interval $(0, 1)$. Bayesian estimates and credible intervals are computed based on draws from the posterior distribution of parameters obtained using an MCMC procedure. Posterior predictive checks, Bayesian standardized residuals and a Bayesian influence measures are considered to check model adequacy, outliers and influential observations. For model selection, we consider the sum of *log*-CPO metric and a Bayesian model selection criteria based on Bayesian mixture modeling. Simulated datasets are used for prior sensitivity analysis and to assess frequentist properties of Bayesian estimates and a real dataset on soil-water retention is analyzed.

Keywords: truncated beta distribution; nonlinear mixed-effects model; Bayesian inference; Bayesian diagnostic; MCMC.

1. Introduction

Truncated data arise when the variable of interest is limited above, below or between specified limits. Thus, in a truncated data, all values outside the truncation limits are omitted and not even a record of the omitted cases is kept. Truncation limits can be fixed known values, resulting from human intervention and controlled by the investigator (see [1]), and they can also be random variables, i.e, the truncation itself can also be the result of a random process, which is not controlled by the investigator but are the outcome of the sampling process and thus, they can be represented by a stochastic process (see [29]).

Truncated linear and nonlinear regression models have been widely studied considering both the parametric and the semi or non-parametric inferential approach. For some work along these lines, see [1, 2, 6, 10]. In contrast to classical methods, Bayesian inference procedures have the advantage of being computationally simple and readily applicable to any sample size.

Therefore, in this work we propose a Bayesian randomly truncated beta nonlinear mixed-effects model which is constructed based on the class of Bayesian nonlinear mixed-effects models [11, 26]. This regression model is constructed assuming the truncated response variable to follow a truncated beta distribution parameterized by a mean pa-

*Corresponding author. Email: carolina_paraiba@yahoo.com.br

rameter and a parameter related to the dispersion of the truncated variable, and by considering its mean parameter to be associated with a nonlinear continuous function of covariates and unknown parameters and with random effects. The mixed-effects model framework allows to account for a possible correlated data structure of the observed data caused by the presence of extra variability due to an unobserved effect. The truncation limits of the response variable are assumed to be random variables themselves and this information is also accounted for in the construction of the proposed model.

The parameters of the proposed Bayesian truncated beta nonlinear mixed-effects models are estimated by generating an MCMC sample from the posterior distribution. Posterior predictive checks [17, 18] and Bayesian residuals are considered to check for model adequacy, outliers and influential observations. The Kullback-Leibler (KL) divergence calibration based on the conditional predictive ordinate (CPO) [7, 24] is used to asses influential cases. Bayesian model selection is performed using the sum of log-CPO, and we also a criterion based on Bayesian mixture models. We provide results based on simulated datasets to assess sensitivity to the choice of prior distribution, to verify the quality of Bayesian estimates, and to investigate the effectiveness of the Bayesian model diagnostic techniques addressed. To illustrate the proposed methodology, we analyze a soil-water retention dataset from the Buriti Vermelho River Basin database of [27].

2. Motivating example: soil-water characteristic curves

We consider soil-water retention data which are useful for constructing soil-water characteristic curves (SWCCs), a graphical representation which describes how a soil store and release water. Usually, the data is obtained experimentally after applying different tensions levels, x, to a given soil sample and observing y, the soil-water content (proportion of water) in the soil sample. Tension levels may be measured in units of atmosphere (atm) pressure.

Since water flow only occurs for soil-water contents measured between the saturation value and the residual value, the observed soil-water content is limited from below by the residual soil-water content and limited from above by the saturated soil-water content, meaning that soil-water retention data is a proportion subjected to truncation. The residual soil-water, a, is the amount of water that cannot be drained from the soil even at high tension values and it is defined as the soil-water content at $x = 15atm$ [28]. The saturated soil-water content, b, is the maximum water content able to be held in the soil before drainage takes place and it is defined as the water content at $0atm$ [28]. These two characteristics of the soil vary among soil samples and are different across soil depths, they shall be treated as being random variables themselves.

Several analytical nonlinear expressions, $\eta(x, \boldsymbol{\beta})$, have been proposed in the literature for representing SWCCs. Among the most widely used SWCCs expressions are Gardner's, van Genuchten's and Fredlund-Xing's [14, 15, 28]. In [21], the authors show that Gardner's, van Genuchten-Mualem's and Fredlund-Xing's expressions, among others, can be derived from a single generic SWCC.

Gardner's expression [15] is given by

$$\eta(x, \boldsymbol{\beta}) = a + \frac{b - a}{1 + \beta_1 x^{\beta_2}}, \tag{1}$$

where β_1 is related to the inverse air entry value of the soil and β_2 is related to the slope of the SWCC.

The van Genuchten-Mualem expression is obtained by applying Mualem's restriction

[23] $\beta_3 = 1 - \frac{1}{\beta_2}$ to β_3 in van Genuchten's expression [28], thus

$$\eta(x, \boldsymbol{\beta}) = a + \frac{b-a}{\left[1 + (\beta_1 x)^{\beta_2}\right]^{1-\frac{1}{\beta_2}}}, \tag{2}$$

where $\beta_1 > 0$ is related to the inverse of the air entry value and $\beta_2 > 1$ since $\beta_3 > 0$.

Fredlund-Xing's expression [14] is written as

$$\eta(x, \boldsymbol{\beta}) = b\left(1 - \frac{\ln\left(1 + \frac{x}{x_r}\right)}{\ln\left(1 + \frac{10^6}{x_r}\right)}\right)\left\{\ln\left[e + \left(\frac{x}{\beta_1}\right)^{\beta_2}\right]\right\}^{-\beta_3}, \tag{3}$$

where β_1 is related to the air entry value, β_2 and β_3 are related to the pore-size distribution, e is the natural number and $x_r = 15atm$ is the soil suction in residual condition.

Accurate estimation of model parameters is a major concern in SWCCs studies since interest rely on the estimated SWCC which has a number of applications. In agriculture, the estimated SWCC may be used to determine field capacity, permanent wilting point and total soil-water availability, which are important characteristics for establishing proper strategies of irrigation management and water balance in the soil.

Traditionally, SWCC are estimated using the nonlinear least squares method. However, under truncation, usual least squares estimates can be biased, inefficient, and inconsistent, which can seriously affect the estimated SWCC. Moreover, soil-water retention data is usually measured by collecting soil samples along a given region and, at each collection site, soil samples are obtained for different soil depths and the measurements on the response variable are made in replicates. The reason why soil samples are collected from different depths at the same collection site is because soil-water retention at different layers of the same soil is subjected to different hydraulic properties and dynamics. Thus, there is evidence of a possible variability in the data due to an unobserved effect of soil depth and it is important to consider a mixed-effect model framework where the random effects account for the unobserved effect that the different soil depth have in soil-water retention. Therefore, we propose SWCCs to be estimated using the alternative models constructed in Section 3.

3. Model formulation

We propose a Bayesian randomly truncated nonlinear beta mixed-effects model to address data subject to both random truncation and unobserved sources of variation. We assume the observed data to be measured in $i = 1, \ldots, M$ groups, with $j = 1, \ldots, N_i$ replicates and $k = 1, \ldots, n_{ij}$ observations each, and that the data is subjected to unobserved variability due to group effect. Truncated beta models are suitable for response variables bounded to an interval (a, b) without the need to consider a rescaled variable to apply the well known beta regression model and its extensions, which are primarily suitable for response variables in the unit interval $(0, 1)$. Moreover, the truncated Beta distribution (5) is a natural alternative for modeling proportions which are truncated to subsets of $(0, 1)$.

In [13], a Bayesian mixed-effects approach was proposed for the well known and widely applicable beta regression model [12]. Although the model is primarily suitable for response variables in the unit interval $(0, 1)$, the authors suggested that the Bayesian beta mixed-effects regression model can be applied to variables bounded to any interval (a, b) by considering the transformation $y^* = (y - a)/(b - a)$. However, as noted in [12], transforming the response variable can make it difficult to interpret model parameters in terms of the original variable. Furthermore, in [8], the author considers it to be more accurate to state that truncation is a characteristic of the probability distribution from which the sample data are drawn and [22] suggests that to account for the truncated nature of the observed data we must consider a truncated probability distribution, which is the part of a distribution that is above, below or between some specified value.

Therefore, we consider the truncated beta distribution as the model for the observed data. We start with the definition and properties of this distribution, followed by the definition of the likelihood and of the prior distribution.

3.1. Properties of truncated beta distribution

Let Y be a $Beta\,(c,d)$ r.v., $c,d>0$. Setting $\mu = c/\,(c+d)$ and $e^{\sigma} = c+d$, we have $E\,(Y) = \mu$ and $Var\,(Y) = \mu\,(1-\mu)/(1+e^{\sigma})$. Thus, $\mu \in (0,1)$ is the mean parameter, and $\sigma \in \Re$ is a parameter related to the dispersion of the r.v. Y and its p.d.f., denoted by $B\,(\mu,\sigma)$, is rewritten as

$$f\,(y) = \frac{y^{\mu e^{\sigma}-1}(1-y)^{(1-\mu)e^{\sigma}-1}}{B\,(\mu e^{\sigma},(1-\mu)\,e^{\sigma})}. \tag{4}$$

If Y is truncated to (a,b), then the p.d.f. of Y given $a<Y<b$, $0<a<b<1$, denoted by $TB\,(\mu,\sigma,a,b)$, is given by

$$f\,(y\,|a<y<b\,) = \frac{y^{\mu e^{\sigma}-1}(1-y)^{(1-\mu)e^{\sigma}-1}}{B\,(b;\mu e^{\sigma},(1-\mu)\,e^{\sigma}) - B\,(a;\mu e^{\sigma},(1-\mu)\,e^{\sigma})}, \tag{5}$$

where $B\,(\kappa,\tau) = \int_0^1 y^{\kappa-1}(1-y)^{\tau-1}dy$ and $B\,(t;\kappa,\tau) = \int_0^t y^{\kappa-1}(1-y)^{\tau-1}dy$ are the beta function and the incomplete beta function, respectively.

The expressions for the expectation and variance of a truncated beta r.v., which will be used for computing posterior predictive checks (Section 4.1) and Bayesian residuals (Section 4.3), are given by

$$E\,(Y\,|a<Y<b\,) = \frac{B\,(a;\mu e^{\sigma}+1,(1-\mu)\,e^{\sigma}) - B\,(b;\mu e^{\sigma}+1,(1-\mu)\,e^{\sigma})}{B\,(a;\mu e^{\sigma},(1-\mu)\,e^{\sigma}) - B\,(b;\mu e^{\sigma},(1-\mu)\,e^{\sigma})}, \tag{6}$$

and

$$\begin{aligned} Var\,(Y\,|a<Y<b\,) &= \frac{B\,(a;\mu e^{\sigma}+2,(1-\mu)\,e^{\sigma}) - B\,(b;\mu e^{\sigma}+2,(1-\mu)\,e^{\sigma})}{B\,(a;\mu e^{\sigma},(1-\mu)\,e^{\sigma}) - B\,(b;\mu e^{\sigma},(1-\mu)\,e^{\sigma})} \\ &- \left[\frac{B\,(a;\mu e^{\sigma}+1,(1-\mu)\,e^{\sigma}) - B\,(b;\mu e^{\sigma}+1,(1-\mu)\,e^{\sigma})}{B\,(a;\mu e^{\sigma},(1-\mu)\,e^{\sigma}) - B\,(b;\mu e^{\sigma},(1-\mu)\,e^{\sigma})}\right]^2, \end{aligned} \tag{7}$$

respectively.

3.2. Nonlinear mixed-effects randomly truncated beta model

Consider Y to be a r.v. of interest subject to random truncation limits A and B, $A<B$. Then, the distribution of the random vector $(Y,A,B)\,|\,(A<Y<B)$, can be defined as the product of the conditional distributions of $Y|\,(A,B,A<Y<B)$, $A|\,(B,A<Y<B) \overset{d}{=} A|B$ and $B|\,(A<Y<B) \overset{d}{=} B$, where $\overset{d}{=}$ denotes equality in distribution. In this paper, we consider $Y\,|(A,B,A<Y<B) \sim TB\,(\mu,\sigma,a,b)$, $A\,|B \sim TB\,(\mu_A,\sigma_A,0,b)$ and $B \sim Beta\,(\mu_B,\sigma_B)$, $\mu_B \in (0,1)$ and $\sigma_B \in \Re$.

Let $\boldsymbol{u}_i$ be a s-dimensional vector of unknown random effects and let $\boldsymbol{U}_{i,jk}$ be a $n_{ij} \times s$ design matrix related to $\boldsymbol{u}_i$. Let $\boldsymbol{u} = \left(\boldsymbol{u}_1^{'}, \ldots, \boldsymbol{u}_M^{'}\right)^{'}$, and consider $\boldsymbol{x} = \left(\boldsymbol{x}_{1,1}^{'}, \ldots, \boldsymbol{x}_{M,N_M}^{'}\right)^{'}$ a vector of p covariates.

Assume $\boldsymbol{Y} \left| (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{A} < \boldsymbol{Y} < \boldsymbol{B}, \boldsymbol{u}) \right.$, $\boldsymbol{A} \left| \boldsymbol{B} \right.$ and $\boldsymbol{B}$ n-dimensional vectors whose elements are $Y_{i,jk} \left| (A_{i,jk}, B_{i,jk}, A_{i,jk} < Y_{i,jk} < B_{i,jk}, \boldsymbol{u}_i) \right.$, $A_{i,jk} \left| B_{i,jk} \right.$ and $B_{i,jk}$ and let $\boldsymbol{y} = \left(\boldsymbol{y}_{1,1}^{'}, \ldots, \boldsymbol{y}_{M,N_M}^{'}\right)^{'}$, $\boldsymbol{a} = \left(\boldsymbol{a}_{1,1}^{'}, \ldots, \boldsymbol{a}_{M,N_M}^{'}\right)^{'}$ and $\boldsymbol{b} = \left(\boldsymbol{b}_{1,1}^{'}, \ldots, \boldsymbol{b}_{M,N_M}^{'}\right)^{'}$ be n-dimensional vectors of observed values of $\boldsymbol{Y} \left| (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{A} < \boldsymbol{Y} < \boldsymbol{B}, \boldsymbol{u}) \right.$, $\boldsymbol{A} \left| \boldsymbol{B} \right.$ and $\boldsymbol{B}$, $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$, $n = \sum_{i=1}^{M} \sum_{j=1}^{N_i} n_{ij}$. Then, the randomly truncated beta nonlinear mixed-effects model is defined as

$$\begin{gathered}
Y_{i,jk} \left| (A_{i,jk}, B_{i,jk}, A_{i,jk} < Y_{i,jk} < B_{i,jk}, \boldsymbol{u}_i; \boldsymbol{\beta}, \sigma) \right. \sim TB\left(\mu_{i,jk}, \sigma, a_{i,jk}, b_{i,jk}\right), \\
\mu_{i,jk} = \eta\left(\boldsymbol{x}_{i,jk}, \boldsymbol{\beta}\right) + \boldsymbol{U}_{i,jk}\boldsymbol{u}_i, \\
A_{i,jk} \left| (B_{i,jk}; \mu_A, \sigma_A) \right. \sim TB\left(\mu_A, \sigma_A, 0, b\right), \\
B_{i,jk} \left| (\mu_B, \sigma_B) \right. \sim Beta\left(\mu_B, \sigma_B\right), \\
\boldsymbol{u}_i \left| \boldsymbol{\Psi} \right. \sim N_s\left(\boldsymbol{0}, \boldsymbol{\Psi}\right), \\
i = 1, \ldots, M,\ j = 1, \ldots, N_i,\ k = 1, \ldots, n_{ij}.
\end{gathered} \tag{8}$$

Model (8) can be reduced for cases where only lower or upper truncation occurs by assuming a distribution for the relevant r.v. of truncation (A, if the data is randomly lower truncated and B if the data is randomly upper truncated) and considering the other limit to be equal to $+\infty$ or $-\infty$ (as convenience). Moreover, model (8) can be reduced to the case where there is no extra variability caused by an unobserved source of variation by assuming $\boldsymbol{u}_i = \boldsymbol{0}$ for all $i = 1, \ldots, M$.

If the response variable is truncated within fixed known truncation limits, then a fixed truncated beta nonlinear mixed-effects model could be written as a particular case of model (8) with $A_{i,jk}$ and $B_{i,jk}$ fixed at $a_{i,jk}$ and $b_{i,jk}$ for each observation $y_{i,jk}$. In this scenario, we would only assume a truncated beta distribution (5) for the responses. It could also be the that $a_{i,jk} = a$ and $b_{i,jk} = b$ for all $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$. If only lower truncation occurs, we set all $b_{i,jk}$ to $+\infty$, and if only upper truncation occurs, we set the $a_{i,jk}$'s to $-\infty$. Finally, this model can be reduced to the case where there is no extra variability caused by an unobserved source of variation by regarding the $\boldsymbol{u}_i$'s as zero.

3.3. Bayesian model

The likelihood function for the observed data $D = (n, \boldsymbol{y}, \boldsymbol{x}, \boldsymbol{a}, \boldsymbol{b})$, given $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma, \boldsymbol{\Psi}, \boldsymbol{u})^{'}$, $\boldsymbol{\omega}_A = (\mu_A, \sigma_A)$ and $\boldsymbol{\omega}_B = (\mu_B, \sigma_B)$ under model (8) is written as

$$\begin{aligned}
&\mathcal{L}\left(D \left| \boldsymbol{\theta}, \boldsymbol{\omega}_A, \boldsymbol{\omega}_B \right.\right) \\
&\propto \prod_{i=1}^{M} \prod_{j=1}^{N_i} \prod_{k=1}^{n_{ij}} \left\{ \frac{y_{i,jk}^{\gamma_{i,jk}(\boldsymbol{\beta},\sigma,\boldsymbol{u})-1} \left(1 - y_{i,jk}\right)^{\lambda_{i,jk}(\boldsymbol{\beta},\sigma,\boldsymbol{u})-1}}{C_{i,jk}\left(\boldsymbol{\beta}, \sigma, \boldsymbol{u}\right)} \right. \\
&\left. \times \frac{a_{i,jk}^{\mu_A e^{\sigma_A}-1} \left(1 - a_{i,jk}\right)^{(1-\mu_A)e^{\sigma_A}-1}}{B\left(b_{i,jk}; \mu_A e^{\sigma_A}, \left(1 - \mu_A\right) e^{\sigma_A}\right)} \frac{b_{i,jk}^{\mu_B e^{\sigma_B}-1} \left(1 - b_{i,jk}\right)^{(1-\mu_B)e^{\sigma_B}-1}}{B\left(\mu_B e^{\sigma_B}, \left(1 - \mu_B\right) e^{\sigma_B}\right)} \right\} \\
&= \mathcal{L}_1\left(D \left| \boldsymbol{\theta} \right.\right) \mathcal{L}_2\left(D \left| \boldsymbol{\omega}_A \right.\right) \mathcal{L}_3\left(D \left| \boldsymbol{\omega}_B \right.\right),
\end{aligned} \tag{9}$$

where

$$\gamma_{i,jk}\left(\boldsymbol{\beta}, \sigma, \boldsymbol{u}\right) = \left[\eta\left(\boldsymbol{x}_{i,jk}, \boldsymbol{\beta}\right) + \boldsymbol{U}_{i,jk}\boldsymbol{u}_i\right] e^{\sigma},$$

$$\lambda_{i,jk}\left(\boldsymbol{\beta},\sigma,\boldsymbol{u}\right)=\left\{1-\left[\eta\left(\boldsymbol{x}_{i,jk},\boldsymbol{\beta}\right)+\boldsymbol{U}_{i,jk}\boldsymbol{u}_i\right]\right\}e^{\sigma}$$

and

$$C_{i,jk}\left(\boldsymbol{\beta},\sigma,\boldsymbol{u}\right)=B\left(b_{i,jk};\gamma_{i,jk}\left(\boldsymbol{\beta},\sigma,\boldsymbol{u}\right),\lambda_{i,jk}\left(\boldsymbol{\beta},\sigma,\boldsymbol{u}\right)\right)-B\left(a_{i,jk};\gamma_{i,jk}\left(\boldsymbol{\beta},\sigma,\boldsymbol{u}\right),\lambda_{i,jk}\left(\boldsymbol{\beta},\sigma,\boldsymbol{u}\right)\right).$$

Note that the random effects model (8) already includes the prior distribution for random effects u:

$$\pi(u \mid \Psi)=\prod_{i=1}^{M}\frac{1}{(2\pi)^{s/2}|\boldsymbol{\Psi}|}\exp\left[-\frac{1}{2}\boldsymbol{u}_i^{'}\boldsymbol{\Psi}^{-1}\boldsymbol{u}_i\right].$$

We assume that the fixed effects and the random effects are independent given σ which is typical for a random effects model. For the considered data, $s=1$ and $U_{i,jk}=1$ for all i and jk, i.e. we can write the mean as $\mu_{i,jk}=\eta\left(\boldsymbol{x}_{i,jk},\boldsymbol{\beta}\right)+\boldsymbol{u}_i$. Therefore, we have $\Psi=\sigma_u$.

We assume that the processes generating the soil-water retention values, the lower truncation limits and the upper truncation limits are independent. This implies independent priors for parameters $\boldsymbol{\theta}$, $\boldsymbol{\omega}_A$ and $\boldsymbol{\omega}_B$, i.e. we can write

$$\pi(\boldsymbol{\theta},\omega_A,\omega_B)=\pi(\boldsymbol{\theta})\pi(\boldsymbol{\omega}_A)\pi(\boldsymbol{\omega}_B).$$

We choose noninformative or weakly informative prior distributions $\pi\left(\boldsymbol{\theta}\right)$, $\pi\left(\boldsymbol{\omega}_A\right)$ and $\pi\left(\boldsymbol{\omega}_B\right)$ for $\boldsymbol{\theta}$, $\boldsymbol{\omega}_A$ and $\boldsymbol{\omega}_B$, respectively. A detailed discussion with sensitivity analysis for the choice of these priors is given in Section 5.1 which suggests the use of the prior distributions given by (17) and (18).

Prior elicitation for the parameters of SWCC is an important issue. Current literature, including prior elicitation, focuses on the parameters of the least squares model which we aim to improve on by considering a more appropriate truncated model which also includes modeling of truncations as random effects. Also, majority of the papers analyze soil samples which differ from the Buriti Vermelho River Basin soil (the soil composition, which affects its hydraulic and dynamic properties influencing soil-water retention). Therefore, to avoid using inaccurate priors and in absence of additional a priori information that has to be specific to the Buriti Vermelho River Basin type of soil analyzed here, we use a priori information to specify the structure of the Bayesian model, namely conditional dependence of parameters, and the particular parametrizations of the distributions (e.g. random effect models for truncation limits), and noninformative or weakly informative priors on the hyperparameters. Once more information is available on this soil, it can be incorporated in the Bayesian model by specifying more informative prior distributions.

4. Bayesian analysis

Using Bayes' theorem and the prior dependence structure of the parameters, the posterior distribution of $\boldsymbol{\theta}$, $\boldsymbol{\omega}_A$ and $\boldsymbol{\omega}_B$ can be written as

$$\begin{aligned}&\pi\left(\boldsymbol{\theta},\boldsymbol{\omega}_A,\boldsymbol{\omega}_B\,|D\right)\propto\pi\left(\boldsymbol{\theta},\boldsymbol{\omega}_A,\boldsymbol{\omega}_B\right)\mathcal{L}\left(D\,|\boldsymbol{\theta},\boldsymbol{\omega}_A,\boldsymbol{\omega}_B\right)\\&=\pi\left(\boldsymbol{\theta}\right)\pi\left(\boldsymbol{\omega}_A\right)\pi\left(\boldsymbol{\omega}_B\right)\mathcal{L}_1\left(D\,|\boldsymbol{\theta}\right)\mathcal{L}_2\left(D\,|\boldsymbol{\omega}_A\right)\mathcal{L}_3\left(D\,|\boldsymbol{\omega}_B\right)\\&=\pi_1\left(\boldsymbol{\theta}\,|D\right)\pi_2\left(\boldsymbol{\omega}_A\,|D\right)\pi_3\left(\boldsymbol{\omega}_B\,|D\right),\end{aligned}\tag{10}$$

Since $\boldsymbol{\theta}$, $\boldsymbol{\omega}_A$ and $\boldsymbol{\omega}_B$ are a posteriori independent, Bayesian analysis for (10) can be partitioned into studying $\pi_1(\boldsymbol{\theta}\,|D)$, $\pi_2(\boldsymbol{\omega}_A\,|D)$ and $\pi_3(\boldsymbol{\omega}_B\,|D)$, which are not available analytically and may be approximated using Markov chain Monte Carlo (MCMC) methods [5]. Moreover, the full conditional distributions of the components of $\boldsymbol{\theta}$, $\boldsymbol{\omega}_A$ and $\boldsymbol{\omega}_B$ are also not available analytically. Therefore, we use a Metropolis-Hastings algorithm to sample from the joint posteriors $\pi_1(\boldsymbol{\theta}\,|D)$, $\pi_2(\boldsymbol{\omega}_A\,|D)$ and $\pi_3(\boldsymbol{\omega}_B\,|D)$ and to perform Bayesian inference on the marginal posterior distributions.

Bayesian estimates of $\boldsymbol{\theta}$, $\boldsymbol{\omega}_A$ and $\boldsymbol{\omega}_B$ are denoted by $\tilde{\boldsymbol{\theta}}$, $\tilde{\boldsymbol{\omega}}_A$ and $\tilde{\boldsymbol{\omega}}_B$ and are computed from the MCMC samples of $\pi_1(\boldsymbol{\theta}\,|D)$, $\pi_2(\boldsymbol{\omega}_A\,|D)$ and $\pi_3(\boldsymbol{\omega}_B\,|D)$, respectively. We compute two types of credible intervals for $\tilde{\boldsymbol{\theta}}$, $\tilde{\boldsymbol{\omega}}_A$ and $\tilde{\boldsymbol{\omega}}_B$: one given by the interquantile range of the posteriors' samples and the other given by the highest posterior density (HPD) interval [5].

The MCMC sample from the posterior distribution $\pi_1(\boldsymbol{\theta}\,|D)$ is obtained using a Metropolis-Hastings within Gibbs type algorithm with each candidate of each model parameter being generated by random walk from a normal distribution with variances given by the elements of the diagonal of the sample Fisher information matrix (the negative Hessian matrix evaluated at the maximum likelihood estimator of the parameters). The MCMC samples of the posterior distributions $\pi_2(\boldsymbol{\omega}_A\,|D)$ and $\pi_3(\boldsymbol{\omega}_B\,|D)$ are obtained using a Metropolis-Hastings type algorithm with candidates generated by random walk from a multivariate normal distribution with covariance matrix given by minus the sample Fisher information matrix for the corresponding parameters. The Gibbs-Metropolis algorithm for each model was written in statistical software R [25].

We note that this Bayesian estimation procedure is applicable to the particular cases of model (8). Under a fixed truncated beta nonlinear mixed-effects models the likelihood function for the the observed data $D = (n, \boldsymbol{y}, \boldsymbol{x}, a, b)$ corresponds to $\mathcal{L}_1(D\,|\boldsymbol{\theta})$ in (9), with $a_{i,jk} = a$ and $b_{i,jk} = b$ for all $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$ and we only use the prior distribution of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma, \boldsymbol{\Psi}, \boldsymbol{u})^{'}$. Under a randomly truncated beta nonlinear model the likelihood corresponds to (9) with $\boldsymbol{U}_{i,jk}\boldsymbol{u}_i = \boldsymbol{0}$ for all $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$ and we only use the prior distributions for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)^{'}$, $\boldsymbol{\omega}_A$ and $\boldsymbol{\omega}_B$. Under a fixed truncated beta nonlinear model the likelihood corresponds to $\mathcal{L}_1(D\,|\boldsymbol{\theta})$ in (9), with $a_{i,jk} = a$, $b_{i,jk} = b$ and $\boldsymbol{U}_{i,jk}\boldsymbol{u}_i = \boldsymbol{0}$ for all $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$ and we use the prior distribution for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma)^{'}$.

4.1. Posterior predictive checks

Posterior predictive checks [17, 18] are aimed to detect differences between the model and observed data by generating replicated datasets, $\boldsymbol{y}^{rep}$, from the posterior predictive distribution and comparing then to the observed dataset by means of discrepancy variables, T, which can be any function of data and model parameters.

For a future observation $\tilde{\boldsymbol{y}}$ of the response variable and assuming $\tilde{\boldsymbol{y}}$ and $\boldsymbol{y}$ to be conditionally independent given $\boldsymbol{\theta}$, the posterior predictive distribution is given by

$$\pi(\tilde{\boldsymbol{y}}\,|D) = \int_{\Theta} \pi(\tilde{\boldsymbol{y}}\,|\boldsymbol{\theta})\,\pi_1(\boldsymbol{\theta}\,|D)\,d\boldsymbol{\theta}. \tag{11}$$

Given an MCMC sample of size L of $\pi_1(\boldsymbol{\theta}\,|D)$, the discrepancy variable may be computed by the following procedure:

(1) for each $\boldsymbol{\theta}_l$, $l = 1, \ldots, L$, sample $\boldsymbol{y}_l^{rep}$ from $\pi(\tilde{\boldsymbol{y}}\,|\boldsymbol{\theta}_l)$;
(2) compute the discrepancy variable $T(\boldsymbol{y})$ for the observed data;
(3) compute the discrepancy variable $T\left(\boldsymbol{y}_l^{rep}\right)$ for each $\boldsymbol{y}_l^{rep}$, $l = 1, \ldots, L$.

If the discrepancy depends on both the data and parameters, then we should obtain $T(\boldsymbol{y}, \boldsymbol{\theta})$ and $T(\boldsymbol{y}^{rep}, \boldsymbol{\theta})$ for each draw of the posterior distribution, $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L$. As suggested by [18], the comparison can be displayed as a histogram of the differences $T(\boldsymbol{y}, \boldsymbol{\theta}) - T(\boldsymbol{y}^{rep}, \boldsymbol{\theta})$, which should contain 0, or as a scatter plot of $T(\boldsymbol{y}, \boldsymbol{\theta})$ against $T(\boldsymbol{y}^{rep}, \boldsymbol{\theta})$, which should be symmetric about a 45° line. In this paper we consider the mean and variance of the response variable, and the model deviance as discrepancy variables.

The posterior predictive p-value is defined as $p = P[T(\boldsymbol{y}^{rep}) \geqslant T(\boldsymbol{y})|y]$, which based on a MCMC sample of the posterior distribution and on the replicated datasets, is estimated as

$$\hat{p} = \frac{\#\{T(\boldsymbol{y}^{rep}) \geqslant T(\boldsymbol{y})\}}{L}. \tag{12}$$

If the posterior predictive p-value is close to 0 or 1, it may suggest that the observed data has an extreme discrepancy variable and the model may be inappropriate. Advice on the interpretation of posterior predictive p-values, which should not be interpreted as a probability that the considered model is true given the data, can be found in [18].

4.2. Prediction

The predicted value, $\tilde{y}_{i,jk}$, of an observation, $y_{i,jk}$, $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$, may be obtained based on an MCMC sample of size L, $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L$, of $\pi_1(\boldsymbol{\theta}|D)$ using the following procedure which sample draws from the posterior predictive distribution (11):

(1) for each $\boldsymbol{\theta}_l$, $l = 1, \ldots, L$, sample $\tilde{y}_{i,jk_l}$ from $\pi(\tilde{y}|\boldsymbol{\theta}_l)$, $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$;
(2) for $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$, set $\tilde{y}_{i,jk}$, the predicted value of the $(i, jk)^{th}$ observation, as

$$\tilde{y}_{i,jk} = \frac{1}{L}\sum_{l=1}^{L} \tilde{y}_{i,jk_l}. \tag{13}$$

4.3. Bayesian residuals

The posterior predictive residuals [30] may be used to check for model adequacy. These residuals are given by $y_{i,jk} - \tilde{y}_{i,jk}$, where $y_{i,jk}$ is the observed value of the $(i, jk)^{th}$ response and $\tilde{y}_{i,jk}$ is its predicted value given by (13). The standardized posterior predictive residual, for $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$ is written as

$$r_{i,jk} = \frac{y_{i,jk} - \tilde{y}_{i,jk}}{\sqrt{Var\left(Y_{i,jk}\,\middle|\,A_{i,jk}, B_{i,jk}, A_{i,jk} < Y_{i,jk} < B_{i,jk}; D, \tilde{\boldsymbol{\theta}}\right)}}, \tag{14}$$

where $\tilde{\boldsymbol{\theta}}$ is the posterior mean of $\boldsymbol{\theta}$, $\tilde{y}_{i,jk}$ is the predicted value of $y_{i,jk}$ as given by expression (13) and $Var(Y_{i,jk}|A_{i,jk}, B_{i,jk}, A_{i,jk} < Y_i < B_{i,jk}; D, \boldsymbol{\theta})$ is the expression of the variance of the truncated beta distribution (7).

Values of $r_{i,jk}$ centered on zero indicate that the model is well fitted to the data.

4.4. Influence

The conditional predictive ordinate (CPO) [7], defined as the predictive density of the $(i,jk)^{th}$ case given the data without the $(i,jk)^{th}$ case, $D_{(-i,jk)}$, for $i = 1,\ldots,M$, $j = 1,\ldots,N_i$, $k = 1,\ldots,n_{ij}$, is given by

$$CPO_{i,jk} = \left[\int_{\Theta} \frac{1}{f\left(y_{i,jk}\left|a_{i,jk}, b_{i,jk}, a_{i,jk} < y_{i,jk} < b_{i,jk}; \boldsymbol{\theta}\right.\right)} \pi_1\left(\boldsymbol{\theta}\left|D\right.\right) d\boldsymbol{\theta}\right]^{-1}.$$

The KL divergence calibration based on CPO uses a Bayesian perspective to case deletion diagnostic and evaluates the influence of a given observation in parameters estimates. The KL divergence, for $i = 1,\ldots,M$, $j = 1,\ldots,N_i$, $k = 1,\ldots,n_{ij}$, is given by

$$\begin{aligned} &K\left(\pi_1\left(\boldsymbol{\theta}\left|D\right.\right), \pi_1\left(\boldsymbol{\theta}|D_{(-i,jk)}\right)\right) \\ &= -\log\left(CPO_{i,jk}\right) + E_{\boldsymbol{\theta}}\left\{\log f\left(y_{i,jk}\left|a_{i,jk}, b_{i,jk} a_{i,jk} < y_{i,jk} < b_{i,jk}; \boldsymbol{\theta}\right.\right)|D\right\}. \end{aligned}$$

Following [24] and [7], the calibration, for $i = 1,\ldots,M$, $j = 1,\ldots,N_i$, $k = 1,\ldots,n_{ij}$, is written as

$$p_{i,jk} = 0.5\left\{1 + \sqrt{1 - \exp\left[-2\hat{K}\left(\pi_1\left(\boldsymbol{\theta}\left|D\right.\right), \pi_1\left(\boldsymbol{\theta}|D_{(-i,jk)}\right)\right)\right]}\right\}, \tag{15}$$

where

$$\begin{aligned} \hat{K}\left(\pi_1\left(\boldsymbol{\theta}\left|D\right.\right), \pi_1\left(\boldsymbol{\theta}|D_{(-i,jk)}\right)\right) &= \log\left\{\frac{1}{L}\sum_{l=1}^{L} \frac{1}{f\left(y_{i,jk}\left|a_{i,jk}, b_{i,jk}, a_{i,jk} < y_{i,jk} < b_{i,jk}; \boldsymbol{\theta}_l\right.\right)}\right\} \\ &+ \frac{1}{L}\sum_{l=1}^{L} \log f\left(y_{i,jk}\left|a_{i,jk}, b_{i,jk}, a_{i,jk} < y_{i,jk} < b_{i,jk}; \boldsymbol{\theta}_l\right.\right), \end{aligned}$$

is the Monte Carlo approximation of the KL divergence between the posterior distribution with the complete data and the posterior distribution with the i^{th} observation deleted based on a MCMC sample of size L of $\pi_1\left(\boldsymbol{\theta}\left|D\right.\right)$.

Values of (15) substantially bigger than 0.5 may indicate that the observation is influential.

4.5. Model selection

The metric known as sum of *log*-CPO is defined as

$$\log\text{-CPO} = \sum_{i=1}^{M}\sum_{j=1}^{N_i}\sum_{k=1}^{n_{ij}} \log\left(CPO_{i,jk}\right), \tag{16}$$

is an estimator of the logarithm of the marginal likelihood and can be used as model selection criterion. The model to be selected is the one providing the larger value of (16) [4].

We also consider a Bayesian model selection methodology based on the Bayesian mixture model framework.

Suppose we have Q competing models $\mathcal{M}_1,\ldots,\mathcal{M}_Q$ and let $\pi_1\left(\boldsymbol{\theta}_q\left|D\right.\right)$ denote the posterior distribution of $\boldsymbol{\theta}$ under each candidate model q, $q = 1,\ldots,Q$. Then, the posterior distribution of $\boldsymbol{\theta}$ may be written as $\pi_1(\boldsymbol{\theta}\,|D) = \sum_{q=1}^{Q} \rho_q \pi_1(\boldsymbol{\theta}_q\,|D)$, where $\boldsymbol{\rho} = (\rho_1, \ldots, \rho_Q)$ are mixing probabilities, with $0 < \rho_q \leq 1$ and $\sum_{q=1}^{Q} \rho_q = 1$, $q = 1, \ldots, Q$.

Associating latent variables $z_{i,jk}$ to each observation $D_{i,jk} = (y_{i,jk}, \boldsymbol{x}_{i,jk}, a_{i,jk}, b_{i,jk})$ such that $\boldsymbol{\theta}\,|(D_{i,jk}; Z_{i,jk} = q) \sim \pi_1(\boldsymbol{\theta}_q\,|D)$ and $P(Z_{i,jk} = q) = \rho_q$, the posterior probability that observation $D_{i,jk}$ was generated from the q^{th} candidate model is defined as

$$P(z_{i,jk} = q\,|D_{i,jk}; \boldsymbol{\theta}, \boldsymbol{\rho}) = \frac{\rho_q \pi_1\left(\boldsymbol{\theta}_q | D_{\,i,jk}\right)}{\sum\limits_{q=1}^{Q} \rho_q \pi_1\left(\boldsymbol{\theta}_q | D_{\,i,jk}\right)}.$$

Consider the prior the distribution of $\boldsymbol{\rho}$ to be given by a Dirichlet distribution with known fixed hyperparameters $\delta_1, \ldots, \delta_Q$, i.e., $\boldsymbol{\rho} \sim D(\delta_1, \ldots, \delta_Q)$. Let $\boldsymbol{z} = \left(\boldsymbol{z}_{1,1}', \ldots, \boldsymbol{z}_{M,N_M}'\right)'$ be a n-dimensional vector of observed values of $\boldsymbol{Z}$, composed by the $Z_{i,jk}$'s, $n = \sum_{i=1}^{M} \sum_{j=1}^{N_i} n_{ij}$, $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$. Then, the posterior probability distribution of $\boldsymbol{\rho}$ given D and $\boldsymbol{z}$, denoted by $\pi(\boldsymbol{\rho}\,|D, \boldsymbol{z})$, is a $D\left(\delta_1^*, \ldots, \delta_Q^*\right)$ distribution, with $\delta_q^* = \delta_q + n_q$, $n_q = \#\{z_{i,jk} = q\}$, $\sum_{q=1}^{Q} n_q = n$, $n = \sum_{i=1}^{M} \sum_{j=1}^{N_i} n_{ij}$, $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$.

Given MCMC samples of each candidate model $\mathcal{M}_q$ with posterior distribution $\pi_1(\boldsymbol{\theta}_q\,|D)$, the following Gibbs sampling procedure may be used to sample from the posterior distribution $\pi(\boldsymbol{\rho}\,|D, \boldsymbol{z})$:

(1) compute the Bayesian estimates $\tilde{\boldsymbol{\theta}}_1, \ldots, \tilde{\boldsymbol{\theta}}_q$ under each candidate model $\mathcal{M}_q$ with posterior distribution $\pi_1(\boldsymbol{\theta}_q\,|D)$, $q = 1, \ldots, Q$;
(2) for each $D_{i,jk}$ compute the posterior densities under each candidate model $\pi_1\left(\tilde{\boldsymbol{\theta}}_q\,|D_{i,jk}\right)$, $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$;
(3) set initial values $\boldsymbol{\rho}^{(0)}$;
(4) for $t = 1, \ldots$:
  - generate $z_{i,jk}^{(t)}$ from $(q = 1, \ldots, Q)$, with
  $$P\left(z_{i,jk}^{(t)} = q\,\middle|D_{i,jk}; \boldsymbol{\theta}, \boldsymbol{\rho}^{(t-1)}\right) = \frac{\rho_q^{(t)} \pi_1\left(\boldsymbol{\theta}_q | D_{\,i,jk}\right)}{\sum\limits_{q=1}^{Q} \rho_q^{(t)} \pi_1\left(\boldsymbol{\theta}_q | D_{\,i,jk}\right)};$$
  - generate $\boldsymbol{\rho}^{(t)}$ from $D\left(\delta_1^*, \ldots, \delta_Q^*\right)$, with $\delta_q^* = \delta_q + n_q$, $n_q = \#\{z_{i,jk} = q\}$, $\sum_{q=1}^{Q} n_q = n$, $n = \sum_{i=1}^{M} \sum_{j=1}^{N_i} n_{ij}$, $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$.

The Bayesian estimate of $\boldsymbol{\rho}$, denoted by $\tilde{\boldsymbol{\rho}}$ and computed from the MCMC sample of $\pi(\boldsymbol{\rho}\,|D, \boldsymbol{z})$, may be interpreted as the posterior probability that the observed data D has been generated from the q^{th} model and so, we set as the selected model the one with higher value of $\tilde{\rho}_q$, $q = 1, \ldots, Q$.

5. Results based on simulated datasets

To illustrate Bayesian estimation and diagnostic under the proposed randomly truncated beta nonlinear mixed-effects model (8) of Section 3, we simulate datasets mimicking the characteristics of soil-water retention data: the response variable, $Y\,|(A, B, A < Y < B)$, is the soil sample water content, which is limited from below by the residual water content, $A\,|B$, the soil-water content of the soil sample measured at $15atm$ and limited from above by the saturated water content, B, the soil-water content of the soil sample measured at

$0atm$. There is only one covariate, x, representing tension levels ranging from $0,01atm$ to $10atm$, $x = (0,01; 0,03; 0,06; 0,1; 0,33; 0,8; 4; 10)$, thus there are $n_{ij} = 8$ observations in the j^{th} replicate of the i^{th} group, with $k = 1, \ldots, n_{ij}$. The datasets are simulated considering that soil samples were collected at three different soil depths, thus $M = 3$, with $i = 1, 2, 3$.

Below we use τ and λ to denote large constants to denote large constants, $IG(c, d)$ to denote an inverse Gamma distribution and $t_+(\nu, \Delta)$ to denote the t distribution with ν degrees of freedom, scale Δ, centered at 0 and truncated to positive semiline. We take $\tau = 10^4$ and $\lambda = 10^2$.

Samples were generated in the statistical software R and MCMC samples of the posterior distributions $\pi_1(\boldsymbol{\theta} \,|D)$, $\pi_2(\boldsymbol{\omega}_A \,|D)$ and $\pi_3(\boldsymbol{\omega}_B \,|D)$ were obtained using the MCMC procedures described in Section 4. Convergence was checked using Geweke's criterion [19], which performs as t-test of two sample means and compares the mean of the first 10% of the chain with the last 50% of the chain. When the to means are not significantly different, Geweke's diagnostic concludes that the posterior samples are drawn from the stationary distribution. More on convergence diagnostics for MCMC algorithms can be found in [3, 9]. Chain sizes were 700000 with burn-in periods of 20000 and a thinning interval of size 200.

5.1. Prior sensitivity analysis

Prior sensitivity analysis is performed by comparing the posterior summaries obtained for a simulated dataset under different choices of non-informative prior distributions for model parameters. It is worthwhile mentioning that, although we present results for only one simulated dataset, we have also performed prior sensitivity analysis for several other datasets simulated in the same setting described here and we have found similar results regarding the performance of the considered priors distributions.

Simulated data is obtained as follows: $B_{i,jk} \,|(\mu_B, \sigma_B) \sim Beta(\mu_B, \sigma_B)$, $A_{i,jk} \,|(B_{i,jk}; \mu_A, \sigma_A) \sim TB(\mu_A, \sigma_A, 0, b_{i,jk})$ and $Y_{i,jk} \,|(A_{i,jk}, B_{i,jk}, A_{i,jk} < Y_{i,jk} < B_{i,jk}, u_i; \boldsymbol{\beta}, \sigma) \sim TB(\mu_{i,jk}, \sigma, a_{i,jk}, b_{i,jk})$, with $\mu_{i,jk} = \eta(x_{i,jk}, \boldsymbol{\beta}) + u_i$ where $\eta(x_{i,jk}, \boldsymbol{\beta})$ is Gardner's SWCC expression given in (1) and $u_i \sim N(0, \sigma_u)$, $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$, $M = 3$, $n_{ij} = 8$, $N_i = 3$, and $n = 72$.

For the remaining parameters, in absence of a priori information, we aim to have non-informative priors. Firstly, we consider the following choice of prior distributions on the parameters:

$$\begin{aligned} &\beta_k \sim IG(\tau^{-1}, \tau^{-1}) \text{ for } k = 1, 2; \quad \sigma \sim IG\left(\tau^{-1}, \tau^{-1}\right), \\ &\mu_k \sim Beta\,(1/2, 1/2) \quad \& \quad \sigma_k \sim N\,(0, \tau)\,, \quad k \in \{A, B\}, \end{aligned} \tag{17}$$

For the model with link function (3), we also set $\beta_3 \sim IG(\tau^{-1}, \tau^{-1})$. Note that the priors for σ, μ_A and μ_B are approximations to the Jeffreys priors as $\tau \to \infty$ [20], and priors for σ_A and σ_B are approximations to the improper prior with a constant density. Since positive parameters β_1, β_2 (and β_3 in model (3)) have a broad interpretation of scale (possibly under a reparameterization), we use an approximation of the Jeffreys prior for a scale parameter, inverse gamma with small shape and scale parameters. For σ_u, we consider two priors: $IG\left(\tau^{-1}, \tau^{-1}\right)$ and the $t_+(\nu, \tau)$ prior. The inverse Gamma prior is a proper approximation to the improper Jeffreys prior for σ_u, and the folded t prior is advocated by [16] which is more efficient at achieving non-informativeness than the inverse-gamma distribution, especially when σ_u can be small. In this paper, we consider the half-Cauchy distribution, which is a special case of the $t_+(\nu, \lambda)$ with $\nu = 1$; thus, we assume the reparameterization $\sigma_u^r = \sqrt{|\sigma_u|}$, and the prior half-Cauchy distribution of σ_u^r has density

$$\pi\left(\sigma_u^r\right) \propto \left[\left(\sigma_u^r\right)^2 + \lambda^2\right]^{-1} \tag{18}$$

with λ a known fixed hyperparameter.

Secondly, we consider the following reparameterization:

$$\begin{aligned} \beta_k^r &= \log\left(\beta_1\right), \quad k = 1, 2 \\ \sigma_u^r &= \log\left(\sigma_u\right), \\ \mu_k^r &= \log\left\{\mu_k/(1-\mu_k)\right\}, \quad k \in \{A, B\}, \end{aligned} \tag{19}$$

and keep σ_A and σ_B unchanged. For this scenario, we set $N\left(0, \tau\right)$ as the prior distribution of each model parameter.

Based on the results presented in Table 1, it is possible to note that the mean and median of all model parameters, other than σ_u, as well as the 95% interquantile and HPD credible intervals, are reasonably similar regardless the choice of prior (among the ones considered). On the other hand, under the prior (17) with a half-Cauchy prior (18) for parameter σ_u, the posterior summaries for this parameter outperform the ones obtained under inverse Gamma prior for σ_u and under the reparametrized model (19), and it is possible to note that using the half-Cauchy prior, parameter σ_u is more accurately estimated. We also note that parameter σ_u seems to be less accurately estimated under the reparameterized model (19). Therefore, for the reminder of this paper, we will work with the prior (17) and a half-Cauchy prior (18) for σ_u. In addition to $\lambda = 100$, we considered other values of $\lambda < 100$ however they did not affect the posterior distribution.

The results shown in Table 2 and Table 3, for parameter μ_A, σ_A, μ_B and σ_B, indicate the two choices of prior scenarios as providing quite similar posterior summaries. Therefore, we choose to work with the first scenario of prior distribution presented in Table 2 and Table 3 for the parameters related with the lower and upper truncation variables, respectively.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

5.2. Frequentist properties of Bayesian estimates

In this section we present results based on simulated datasets to assess the frequentist properties of Bayesian estimates obtained for the the randomly truncated beta nonlinear mixed-effects model (8) of Section 3.

Simulated datasets are obtained as follows: $B_{i,jk}\left|\left(\mu_B, \sigma_B\right)\right. \sim Beta\left(\mu_B, \sigma_B\right)$, $A_{i,jk}\left|\left(B_{i,jk}; \mu_A, \sigma_A\right)\right. \sim TB\left(\mu_A, \sigma_A, 0, b_{i,jk}\right)$ and $Y_{i,jk}\left|\left(A_{i,jk}, B_{i,jk}, A_{i,jk} < Y_{i,jk} < B_{i,jk}, u_i; \boldsymbol{\beta}, \sigma\right)\right. \sim TB\left(\mu_{i,jk}, \sigma, a_{i,jk}, b_{i,jk}\right)$, with $\mu_{i,jk} = \eta\left(x_{i,jk}, \boldsymbol{\beta}\right) + u_i$ where $\eta\left(x_{i,jk}, \boldsymbol{\beta}\right)$ is Gardner's expression given in (1) and $u_i \sim N\left(0, \sigma_u\right)$, $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$, $M = 3$, $n_{ij} = 8$. Since soil-water retention data are measure in replicates, we consider $N_i = \{1; 3; 30\}$ replicates, which amounts for $n = \{24; 72; 720\}$ sample sizes in the corresponding group. A total of 200 datasets were simulated for each sample size.

We fit the nonlinear truncated beta model (8) with $\eta\left(x_{i,jk}, \boldsymbol{\beta}\right)$ being Gardner's SWCC

(1) with the prior distributions for the model parameters given by (17) and (18).

Simulation results are displayed in Table 7, where the Posterior mean column refers to the mean of the posterior means obtained for each simulated sample; the Bias-m column presents the mean of the bias of the posterior means; the MSE-m column provides the mean of the mean square error (MSE) computed for all posterior means; the Posterior median column presents the median of the posterior medians obtained for each simulated sample; the Bias-md column presents the median of the bias of the posterior medians; the MSE-md column provides the median of the mean square error (MSE) computed for all posterior medians; the Interquantile and the HPD columns provide the estimated coverage probability of 95% interquantile and HPD credibility intervals, respectively. The column Acceptance rate shows the acceptance rate of candidates for each parameter.

Based on the obtained results (Table 7), it can be argued that the Bayesian estimation procedure provided quite precise estimates for the unknown fixed-effects parameters of the nonlinear function $\eta\left(x_{i,jk}, \boldsymbol{\beta}\right)$, β_1 and β_2, related to the mean of the responses, $\mu_{i,jk}$. Parameter σ, related with the dispersion of the responses, is also precisely estimated. Moreover, for β_1, β_2 and σ, both the bias and MSE are very small and the estimated coverage probability of interquantile and HPD credible interval correctly approaches the nominal expected of 95%. On the other hand, we note that the Bayesian estimate of the fixed-effect parameter σ_u associated with the random effects is not as precise as it would be desired. This parameters seems to be underestimated and its estimated standard deviation also seems to be small, causing the estimated credible intervals to be small and, consequently the estimated coverage probabilities for this parameter do not approach the nominal one of 95%. Nevertheless, the Bayesian estimates of the random effects u_1, u_2 and u_2 seem to be satisfactory and their estimated coverage probabilities approach the expected nominal of 95%. Acceptance rates for model parameters across simulated datasets are kept in a reasonable range between 0.3 and 0.7, with smaller values (0.1-0.2) for σ_u which is the most difficult parameter to estimate.

[Table 4 about here.]

5.3. Posterior predictive checks

To asses if posterior predictive checks are capable of identifying model misspecification in the regression structure assumed for the mean parameter, we consider simulated datasets assuming $Y_{i,jk}\left|(a < Y_{i,jk} < b, u_i) \sim TB\left(\mu_{i,jk}, \sigma, a, b\right)\right.$, where $a = 0.31$, $b = 0.55$, and $\mu_{i,jk} = \eta\left(x_{i,jk}, \boldsymbol{\beta}\right) + u_i$, where $\eta\left(x_{i,jk}, \boldsymbol{\beta}\right)$ is Fredlund-Xing's SWCC expression (3) and $u_i \sim N\left(0, \sigma_u\right)$, $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$, $M = 3$, $n_{ij} = 8$, $N_i = 3$. Parameter values are set at $\beta_1 = 0.05$, $\beta_2 = 2$, $\beta_3 = 0.3$, $\sigma = 4$ and $\sigma_u = 0.15$. After simulating the data, the model is fitted considering $Y_{i,jk}\left|(a < Y_{i,jk} < b, u_i) \sim TB\left(\eta\left(x_{i,jk}, \boldsymbol{\beta}\right) + u_i, \sigma, a, b\right)\right.$, with $\eta\left(x_{i,jk}, \boldsymbol{\beta}\right)$ Gardner's SWCC expression (1), with priors (17) and (18). A total of 100 datasets were simulated.

Figure 1 shows the histograms of estimated posterior predictive p-values (12) computed across each simulated dataset using the mean and variance of the response variable, and the model deviance as discrepancy variables. For all three considered discrepancies, we note that the mean, variance and model deviance of the replicated datasets coincide with the obtained values of mean, variance and model deviance of the observed data. This is also supported by the median values of the estimated posterior predictive p-values, which were 0.47, 0.46 and 0.50 for the mean, variance and model deviance, respectively. There is also a high variability among estimated posterior predictive p-values based on the variance discrepancy, which could indicate that this choice of discrepancy variable is not a good choice for model checking under the proposed model and caution should be taking when analyzing it.

Figure 2 illustrate the histograms of the discrepancy variables computed for one simulated dataset. The histograms indicate that the mean, variance and model deviance discrepancy variables are not capable of identifying a significant disagreement between the observed (simulated) data and the replicated datasets.

Therefore, posterior predictive checks do not seem to be capable of correctly identifying lack of fit in the proposed scenario, i.e, we are not able to detect that there is misspecification of the function $\eta\left(x_{i,jk}, \boldsymbol{\beta}\right)$. However, we note that, as mentioned in Section 2, both (1) and (3) may be derived from a common generic expression proposed by [21] and thus, these expressions may be be sufficiently similar as not to produce significantly different results for the observed and replicated data. We note that similar results are obtained when the other SWCCs are compared to each other.

[Figure 1 about here.]

[Figure 2 about here.]

5.4. Bayesian diagnostics

For outlier and influence diagnostic we take one dataset simulated from $Y_{i,jk}\left|(A_{i,jk}, B_{i,jk}, A_{i,jk} < Y_{i,jk} < B_{i,jk}, u_i; \boldsymbol{\beta}, \sigma\right) \sim TB\left(\mu_{i,jk}, \sigma, A_{i,jk}, B_{i,jk}\right)$, with $\mu_{i,jk} = \eta\left(x_{i,jk}, \boldsymbol{\beta}\right) + u_i$, where $\eta\left(x_{i,jk}, \boldsymbol{\beta}\right)$ is Gardner's SWCC (1) and $u_i \sim N\left(0, \sigma_u\right)$, with $B_{i,jk}\left|(\mu_B, \sigma_B\right) \sim Beta\left(\mu_B, \sigma_B\right)$ the upper truncation variable and $A_{i,jk}\left|(B_{i,jk}; \mu_A, \sigma_A\right) \sim TB\left(\mu_A, \sigma_A, 0, b_{i,jk}\right)$ the lower truncation variable, $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$, $M = 3$, $n_{ij} = 8$, $N_i = 3$. The true values of model parameter are: $\beta_1 = 2.25$, $\beta_2 = 0.5$, $\sigma = 6$, $\sigma_u = 0.1$, $\mu_A = 0.25$, $\sigma_A = 5.2$, $\mu_B = 0.6$ and $\sigma_B = 4.6$.

The perturbation in the dataset was induced by transforming observations $\#1, 15$ and $\#2, 16$, into atypical ones by recalculating $y_{1,15}$ and $y_{2,16}$ by adding 2.5 times the standard deviation of y to their original values. The main idea is to assess the effectiveness of the standardized posterior predictive residuals and of the calibration measure and to illustrate their ability to detect influential outliers observation when the randomly truncated beta nonlinear mixed-effects model is fitted to a dataset.

We fit the nonlinear truncated beta model (8) with $\eta\left(x_{i,jk}, \boldsymbol{\beta}\right)$ being Gardner's SWCC (1) with the prior distributions for the model parameters given by (17) and (18).

Figure 3(a) show that both observation $\#1, 15$ and $\#2, 16$ are identified as outliers by the standardized standardized posterior predictive residuals. From Figure 3(b) it is possible to see that KL divergence calibration measure correctly identifies the disturbed cases, $\#1, 15$ and $\#2, 16$, as influential observations.

Although we present results for only one simulated dataset, we note that all metrics were computed to several other data simulated in the same setting for which we have observed these same pattern of behavior.

[Figure 3 about here.]

5.5. Bayesian model selection

To illustrate the performance of the Bayesian model selection metrics, we consider simulated datasets generated by assuming fixed known truncation limits and $Y_{i,jk}\left|(a < Y_{i,jk} < b, u_i\right) \sim BT\left(\mu_{i,jk}, \sigma, a, b\right)$, $i = 1, \ldots, M$, $j = 1, \ldots, N_i$, $k = 1, \ldots, n_{ij}$, $M = 3$, $n_{ij} = 8$, $N_i = 3$. We set $a = 0.3$, $b = 0.6$ and assume $\mu_{i,jk} = \eta\left(x_{i,jk}, \boldsymbol{\beta}\right) + u_i$, where $\eta\left(x_{i,jk}, \boldsymbol{\beta}\right)$ is Gardner's SWCC expression (1) and $u_i \sim N\left(0, \sigma_u\right)$, $i = 1, 2, 3$. Parameter values were set as $\beta_1 = 2.25$, $\beta_2 = 0.5$, $\sigma = 6$ and $\sigma_u = 0.1$. A total of 100 datasets were simulated.

After generating each dataset, two regression models were fitted: $Y_{i,jk}\,|(a < Y_{i,jk} < b, u_i) \sim BT\left(\eta\left(x_{i,jk}, \boldsymbol{\beta}\right) + u_i, \sigma, a, b\right)$, with $\eta\left(x_{i,jk}, \boldsymbol{\beta}\right)$ Gardner's SWCC (1), henceforth model M_0; $Y_{i,jk}\,|(a < Y_{i,jk} < b, u_i) \sim BT\left(\eta\left(x_{i,jk}, \boldsymbol{\beta}\right) + u_i, \sigma, a, b\right)$, with $\eta\left(x_{i,jk}, \boldsymbol{\beta}\right)$ van Genuchten-Mualem's SWCC expression (2), henceforth model M_1.

Under model M_0 prior distributions of parameters are given by (17) and (18). Under model M_1, we assume the same prior distributions for parameters β_1, σ, σ_u, μ_A, μ_B, σ_A, σ_B, but since $\beta_2 > 1$ we consider an inverse gamma prior truncated at 1, which we denote by $TIG\left(\tau^{-1}, \tau^{-1}, 1 + \infty\right)$.

In Figure 4(a) we present the histogram of the of $\tilde{\boldsymbol{\rho}} = \left(\tilde{\rho}_{M_0}, \tilde{\rho}_{M_1}\right)$ obtained across simulated datasets. Similarly, Figure 4(b) shows the histogram of log-CPO values obtained for both model M_0 and M_1 across all simulated datasets. Table 5 displays the median value of the Bayesian estimates of the vector of mixing probabilities $\tilde{\boldsymbol{\rho}}$, the median value of the sum of log-CPO, and the proportion of samples for which the true model M_0 was selected according to both model selection criteria was 0.57 across the 100 simulated datasets. From the obtained results, it is possible to notice that the truncated beta van Genuchten-Mualem mixed-effects model is correctly indicated as the better fit for the dataset by the two criteria. That is, the simulation results indicates that the selection criteria are able to identify the correct model (from which the data were generated) as the most appropriate for the dataset. We note, however, that the values of the mixing probabilities for the two models are reasonably close. This is due to the fact that both Gardner's and van Genuchten-Mualem's equation can be derived from a more general expression, being very similar to each other and thus causing the proportion of of samples for which the true model was selected to be arguably small.

[Table 5 about here.]

[Figure 4 about here.]

6. Application

We apply the proposed methodology ta a soil-water retention dataset from the Buriti Vermelho River Basin, located in the eastern part of the Federal District in Brazil [27]. In the Buriti Vermelho River Basin database, soil samples were collected at depths of $0 - 5cm$, $15 - 20cm$, and $60 - 65cm$ and measured in eight tension levels $x = (0.01, 0.03, 0.06, 0.10, 0.33, 0.80, 4, 10)$ in units of atmosphere pressure (atm). At each tension level, measurements were taken in triplicates. Therefore, the dataset features $M = 3$ depth levels (groups), $N_i = 3$ replicates, and $n_{ij} = 8$ observations per replicate per depth level, and thus $n = \sum_{i=1}^{M} \sum_{j=1}^{N_i} n_{ij} = 72$, with $i = 1, 2, 3$ denoting the depth levels, $j = 1, 2, 3$ denoting the replicates and $k = 1, \ldots, = 8$ denoting the observations per replicate per depth level. For each soil sample, the residual water content was measured at $x = 15atm$ [28] and the saturated soil-water content was measured at $0atm$ [28].

Different structures of SWCCs were taken into account by fitting three different randomly truncated beta nonlinear mixed-effects models (8), thus assuming the randomly truncated responses $Y_{i,jk}\,|(A_{i,jk}, B_{i,jk}, A_{i,jk} < Y_{i,jk} < B_{i,jk}, u_i; \boldsymbol{\beta}, \sigma)$ to follow a $TB\left(\eta\left(\boldsymbol{x}_{i,jk}, \boldsymbol{\beta}\right) + u_i, \sigma, a_{i,jk}, b_{i,jk}\right)$ distribution with random truncation limits $B_{i,jk}\,|(\mu_B, \sigma_B) \sim Beta\left(\mu_B, \sigma_B\right)$ and $A_{i,jk}\,|(B_{i,jk}; \mu_A, \sigma_A) \sim TB\left(\mu_A, \sigma_A, 0, b_{i,jk}\right)$, and considering the three SWCCs expressions presented in Section 2: Gardner's (1), van Genuchten-Mualem's (2) and Fredlund-Xing's (3). Therefore, three models, namely, the Bayesian randomly truncated beta Gardner mixed-effects model (M_1), the Bayesian randomly truncated beta van Genuchten mixed-effects model (M_2) and the Bayesian randomly truncated beta Fredlund-Xing mixed-effects model (M_3).

Under models M_1 and M_3 the prior distributions of parameters are given by (17) and (18). Under model M_2, the prior distribution of $\beta_2 > 1$ is $TN_1(0, \tau)$ - normal distribution truncated at 1, and and the other priors are unchanged.

Values of log-CPO and $\tilde{\boldsymbol{\rho}}$ where 689.65 and 0.48 for model M_1, 648.2 0.46 for model M_2, and 506.72 and 0.06 for model M_3. Therefore, both selection criteria indicate model M_1 as the best fit for the dataset, with model M_2 being marginally second best. Posterior summaries of model M_1 parameters are shown in Table 6 where all fixed-effects parameters are indicated as significantly different from 0 at 95% credibility level. The random effect u_3, related to the unobserved variability associated with soil depth $60 - 65cm$ is the only one indicated as significantly different from 0 at 95% credibility level.

Computed estimated posterior predictive p-values (12) for discrepancies based on the mean and variance of the response variable and model deviance for the selected model M_1 were 0.46, 0.44 and 0.5. Thus, based on estimated posterior predictive p-values, it could be argued that model M_1 seem to be a plausible modeling choices for these dataset, as no extreme $\hat{p}$ were observed, which indicates that the pattern of the observed data is more likely to occur under these models. This conclusion is also attested by the histograms presented in Figure 5. From Figure 5(a) it can be observed that the peak of the histogram of the mean discrepancy coincide with the mean value of the responses. Figure 5(b) shows that the variance of the observed responses is at the peak of the variance discrepancy histograms. Finally, Figure 5(c) reveals that the replicated data is in agreement with the observed data since the value zero is contained in the histogram.

In Figure 6(b) the standardized posterior predictive residuals do not indicate any observation as outlier. Also, the standardized residuals are randomly clustered around zero, which gives evidence of a good fit of the selected model 1. In Figure 6(c) the calibration values do not indicate any case as influential.

Figure 6(a) shows the observed values of the responses and the estimated SWCCs at each soil depth, where it is possible to notice that soil-water retention tends to be higher at depth $15 - 20cm$ (dashed lines with circles) and at depth $60 - 65cm$ (dotted lines with triangles). For both depth $15 - 20cm$ and $60 - 65cm$ we can see there is heterogeneity in soil-water content measures among the replicates at each level. We also notice that data from soil depth $0 - 5cm$ (dotted-dashed lines with diamonds) seems to present lower heterogeneity and lower values of water retention. We also notice that Gardner's equation estimated using the proposed methodology provides a good fit for the representation of the relationship between soil-water content and matric potential for the analyzed soil profile.

Figure 7 shows the contour plot of the joint posterior distribution of β_1 and β_2, and the marginal posterior distributions of β_1, β_2, σ and σ_u. The contour plot of the joint posterior distribution of μ_A and σ_A and their marginal posterior distributions are shown in Figure 8 and the contour plot of the joint posterior distribution of μ_B and σ_B and their marginal posterior distributions are shown in Figure 9. The joint posterior contour plots indicate *a posteriori* dependence between the model parameters β_1 and β_2, and that each pair of the parameters of the truncated beta distribution, μ_k and σ_k for $k \in \{A, B\}$, are approximately *a posteriori* uncorrelated.

[Table 6 about here.]

[Figure 5 about here.]

[Figure 6 about here.]

[Figure 7 about here.]

[Figure 8 about here.]

[Figure 9 about here.]

7. Conclusions

In this work we investigated a class of Bayesian truncated nonlinear beta regression model to study datasets where the response variable is characterized by truncation, a situation that can be observed in different practical applications of scientific knowledge. The truncated nature of the data has been incorporated in the regression model by assuming the response variable to follow a truncated beta probability distribution and the truncation limits have been regarded as random variables themselves and this information was also accounted for in the construction of the proposed model. Moreover, information contained in covariates were incorporated into the model by assuming a nonlinear regression structure for the the location parameter of the truncated responses. The proposed model also accounts for a possible correlated data structure of the observed data caused by the presence of extra variability due to an unobserved effect. The proposed Bayesian model allows incorporation of expert prior knowledge in the analysis to obtain not only the estimates of the unknown parameters but also their uncertainty.

A Bayesian estimation procedure to compute model parameter estimates was successfully applied to the proposed models and simulation results indicated good properties of Bayesian parameter estimates obtained for the proposed truncated nonlinear model. Moreover, diagnostic tools were used to check for model misspecification and to detect outliers and influential observations.

The proposed methodology was applied to construct soil-water characteristic curves, which are important to study the relationship between soil and water, a physical phenomenon that affects soil use in many different purposes and which is of fundamental interest in hydraulic resources management and agriculture. Three nonlinear regression models were compared on the basis of their fit to the data. According to the sum of log-CPO criterion, Gardner's expression provided a better fit to the dataset and the selected models was shown to be well fitted to the data by applying posterior predictive checks, Bayesian residuals and a Bayesian influence diagnostic metric. We also performed model selection between the three competing models by considering a mixture model framework. The posterior distribution of the latent allocation variable for this mixture was successfully used to decide which model fits the data best.

Acknowledgments

The authors would like to thank Lineu Neiva Rodrigues (Embrapa, Cerrados) and Aline de Holanda Nunes Maia (Empraba, Meio Ambiente) for proving the Buruti Vermelho Riven Basin database and for valuable discussion about the experimental data. The research of Carolina Costa Mota Paraíba is supported by the Brazilian organization CAPES.

The algorithms are available upon request to the corresponding author.

References

[1] A'Hearn, A. (2004). A restricted maximum likelihood estimator for truncated height samples. *Economics and Human Biology*, **2**, 5–19.

[2] Bragato, P. L. (2004). Regression analysis with truncated samples and its application to ground-motion attenuation studies. *Bulletin of the Seismological Society of America*, **94**(4), 1369–1378.

[3] Brooks, S., Gelman, A. & Jones, G. L. Meng, X. L. (2011). *Handbook of Markov Chain Monte Carlo*. Handbooks of Modern Statistical Methods Periodical. hapman & Hall/CRC, Boca Raton, first edition.

[4] Carlin, B. P. & Louis, T. A. (2009). *Bayesian methods for data analysis*. Chapman and Hall/CRC, Boca Raton, third edition.

[5] Chen, M. H., Shao, Q. M. & Ibrahim, J. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer, New York.

[6] Chen, S. & Zhou, X. (2012). Semiparametric estimation of a truncated regression model. *Journal of Econometrics*, **167**, 297–304.

[7] Cho, H., Ibrahim, J. G., Sinha, D. & Zhu, H. (2009). Bayesian case influence diagnostics for survival models. *Biometrics*, **65**(1), 116–124.

[8] Cohen, A. C. (1991). *Truncated and censored samples: theory and applications*. Marcel Dekker, New York.

[9] Cowles, M. K. & Carlin, B. P. (1996). Markov chain monte carlo convergence diagnostics: A comparative review. *Journal of the American Statistical Association*, **91**(434), 883–904.

[10] de Uña Álvarez, J., Liang, H.-Y. & Rodríguez-Casal, A. (2010). Nonlinear wavelet estimator of the regression function under left-truncated dependent data. *Journal of Nonparametric Statistics*, **22**(3), 319–344.

[11] Dey, D. K., Chen, M. H. & Chang, H. (1997). Bayesian approach for nonlinear random effects models. *Biometrics*, **53**(4), 1239–1252.

[12] Ferrari, S. L. P. & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, **31**(7), 799–815.

[13] Figueroa-Zúñiga, J. I. F., Arellano-Valle, R. B. & Ferrari, S. L. P. (2013). Mixed beta regression: A bayesian perspective. *Computational Statistics and Data Analysis*, **61**, 137–147.

[14] Fredlund, D. G. & Xing, A. (1994). Equations for the soil-water characteristic curve. *Canadian Geotechnical Journal*, **31**(3), 521–532.

[15] Gardner, W. R. (1958). Some steady state solutions of unsaturated moisture flow equations with application to evaporation from water table. *Soil Science*, **85**, 228–232.

[16] Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, **1**(3), 515–533.

[17] Gelman, A., Goegebeur, Y., Tuerlinckx, F. & Van Mechelen, I. (2000). Diagnostic checks for discrete data regression models using posterior predictive simulations. *Journal of the Royal Statistical Society, Series C*, **49**(2), 247–268.

[18] Gelman, A., Carlin, J., Stern, H. & Rubin, D. B. (2003). *Bayesian data analysis*. Chapman and Hall/CRC Texts in Statistical Science, London, second edition.

[19] Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to calculating posterior moments. In J. M. Bernardo, J. Berger, A. P. Dawid, & J. F. M. Smith, editors, *Bayesian Statistics 4*. Oxford University Press, Oxford.

[20] Jeffreys, H. (1998). *The theory of probability*. Oxford Classic Texts in the Physical Sciences. OUP, Oxford, UK, third edition.

[21] Leong, E. C. & Rahardjo, H. (1997). Review of soil water characteristic curve equations. *Journal of Geotechnical and Geoenvironmental Engineering*, **123**(12), 1106–1117.

[22] Maddala, G. S. (1983). *Limited dependent and qualitative variables in econometrics*. Cambridge, New York.

[23] Mualem, Y. (1976). A new model for predicting the hydraulic conductivity of unsaturated porous media. *Water Resources Research*, **12**, 593–622.

[24] Peng, F. & Dey, D. (1995). Bayesian analysis of outlier problems using divergence measures. *The Canadian Journal of Statistics*, **23**(2), 199–213.

[25] R Development Core Team (2009). *R: A Language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

[26] Racine-Poon, A. (1985). A bayesian approach to nonlinear random effects models. *Biometrics*, **41**(4), 1015–1023.

[27] Rodrigues, L. N. & Maia, A. H. N. (2011). Pedotransfer functions for estimating the saturated hydraulic conductivity and the residual and saturated soil-water content for soils from a watershed located at the brazilian cerrado (in Portuguese). In *XIX Brazilian symposium on water resources*.

[28] van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal*, **44**, 892–898.

[29] Wang, M. C. (1989). A semiparametric model for randomly truncated data. *Journal of the American Statistical Association*, **84**(407), 742–748.

[30] Yan, G. & Sedransk, J. (2010). A note on bayesian residuals as a hierarchical model diagnostic technique. *Statistical Papers*, **51**(1), 1–10.

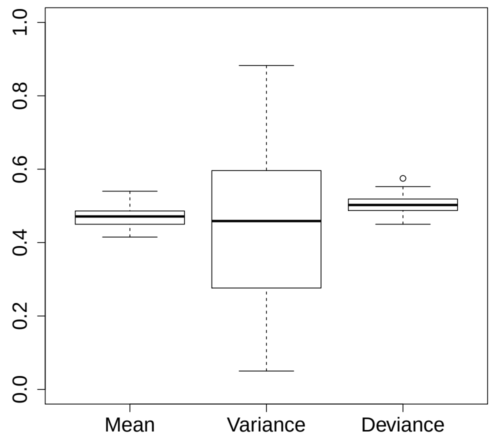


Figure 1. Simulated data: histogram of estimated posterior predictive p-values based on the mean, variance and deviance discrepancy variables.

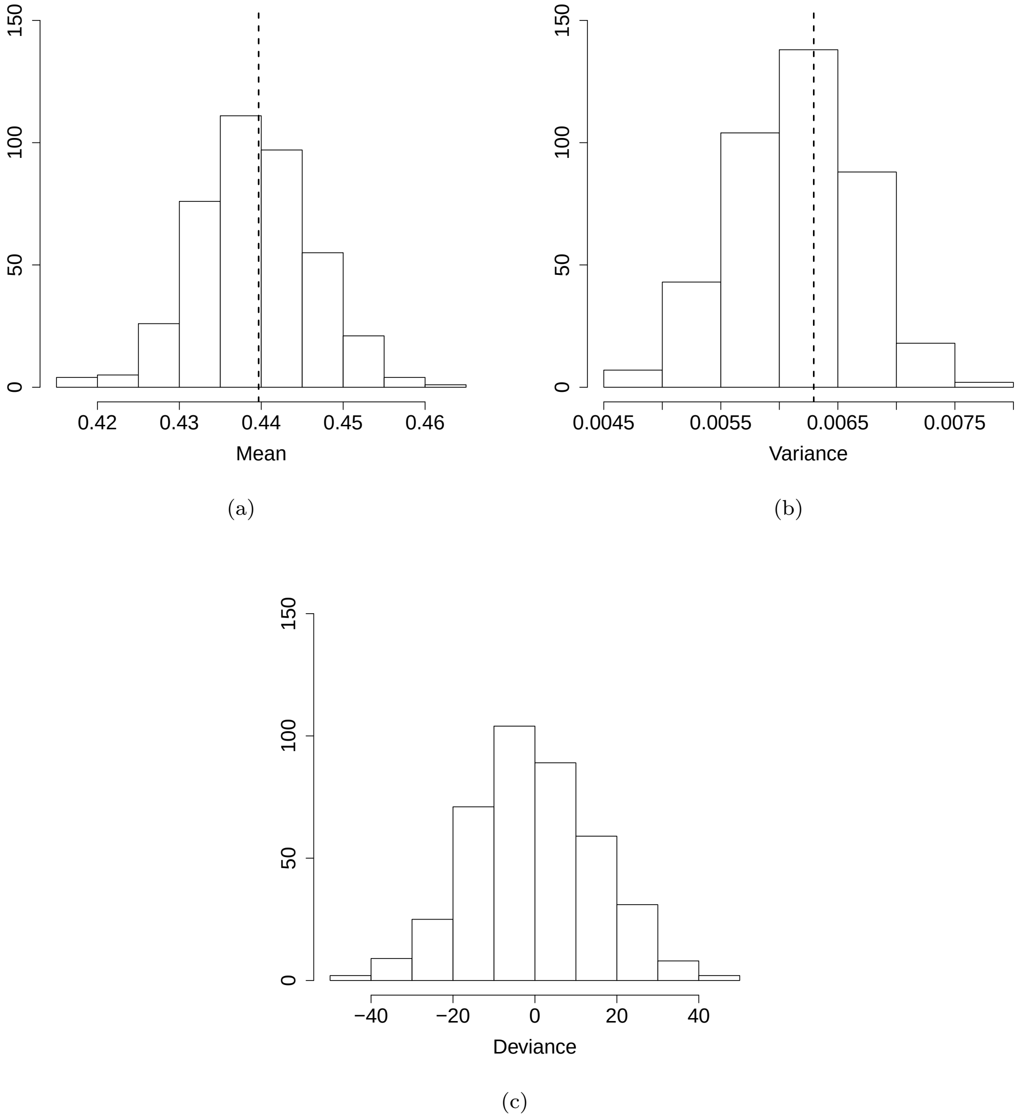
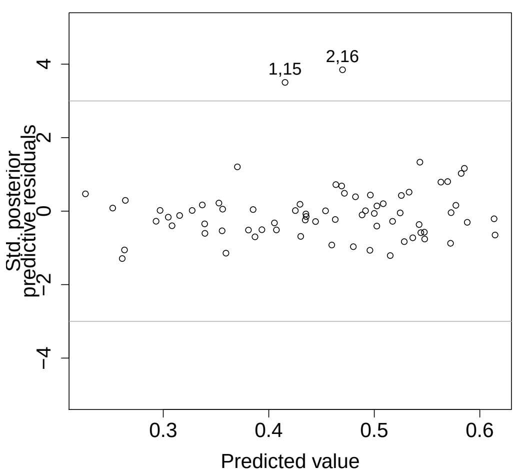
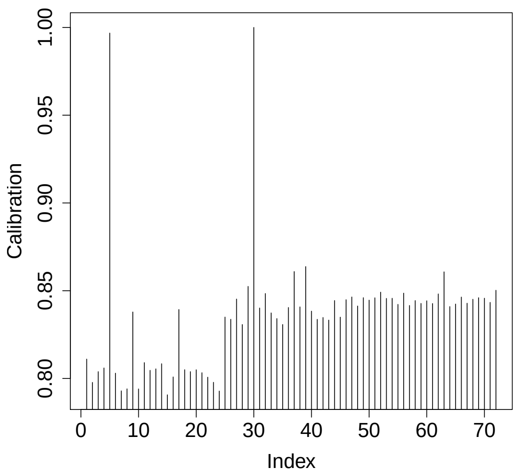


Figure 2. Simulated dataset: histogram of posterior predictive checks based on the mean discrepancy (a), variance discrepancy (b) and model deviance discrepancy (c).

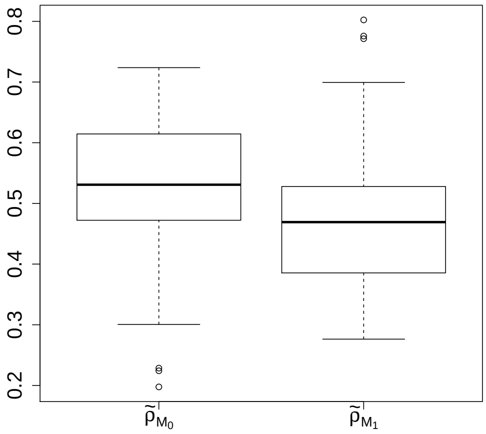


(a)

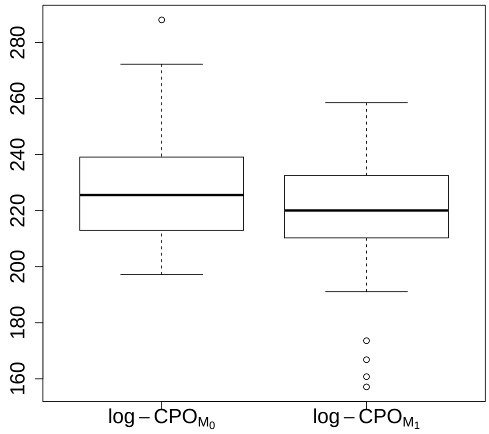


(b)

Figure 3. Simulated data: (a) standardized posterior predictive residuals; (b) calibration measure.

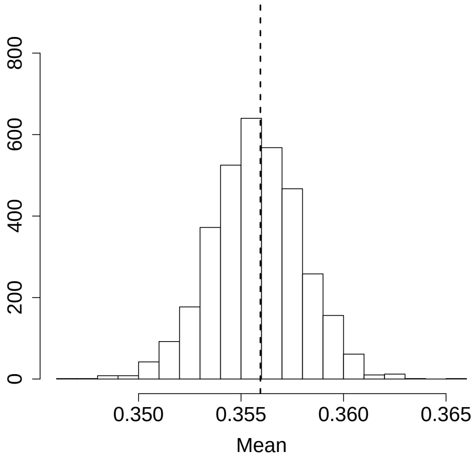


(a)

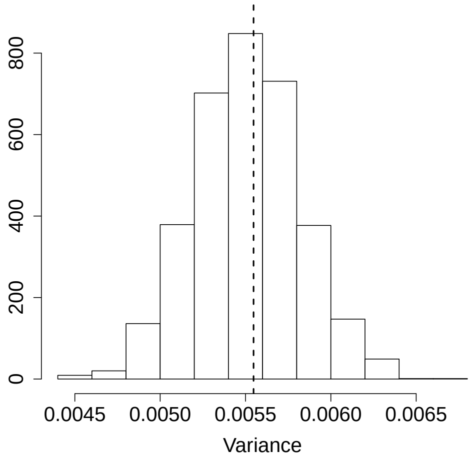


(b)

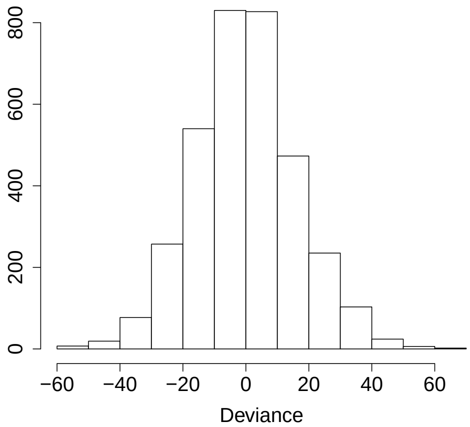
Figure 4. Simulated data: (a) histogram of the Bayesian estimates of the vector of mixing probabilities $\tilde{\boldsymbol{\rho}} = (\tilde{\rho}_{M_0}, \tilde{\rho}_{M_1})$; (b) histogram of log-CPO values ($\log - CPO_{M_0}$ denotes the sum of log-CPO of model M_0 and $\log - CPO_{M_1}$ denotes the sum of log-CPO of model M_1).



(a)

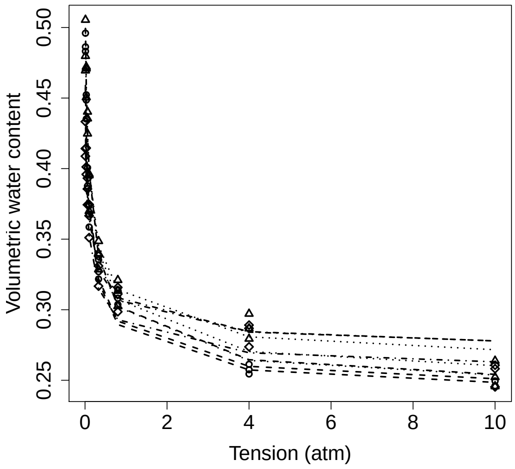


(b)

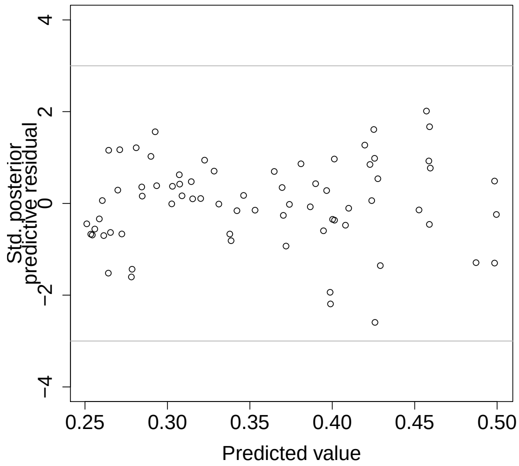


(c)

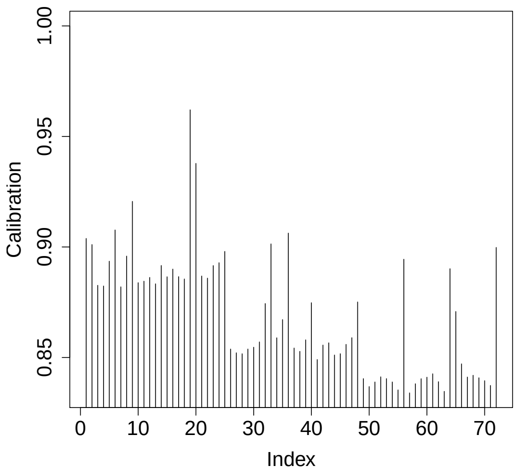
Figure 5. (a) Posterior predictive checks based on the mean discrepancy; (b) posterior predictive checks based on the variance discrepancy; (c) posterior predictive checks based on model deviance.



(a)



(b)



(c)

Figure 6. (a) Estimated Gardner's SWCC; (a) standardized posterior predictive residuals; (b) calibration measure.

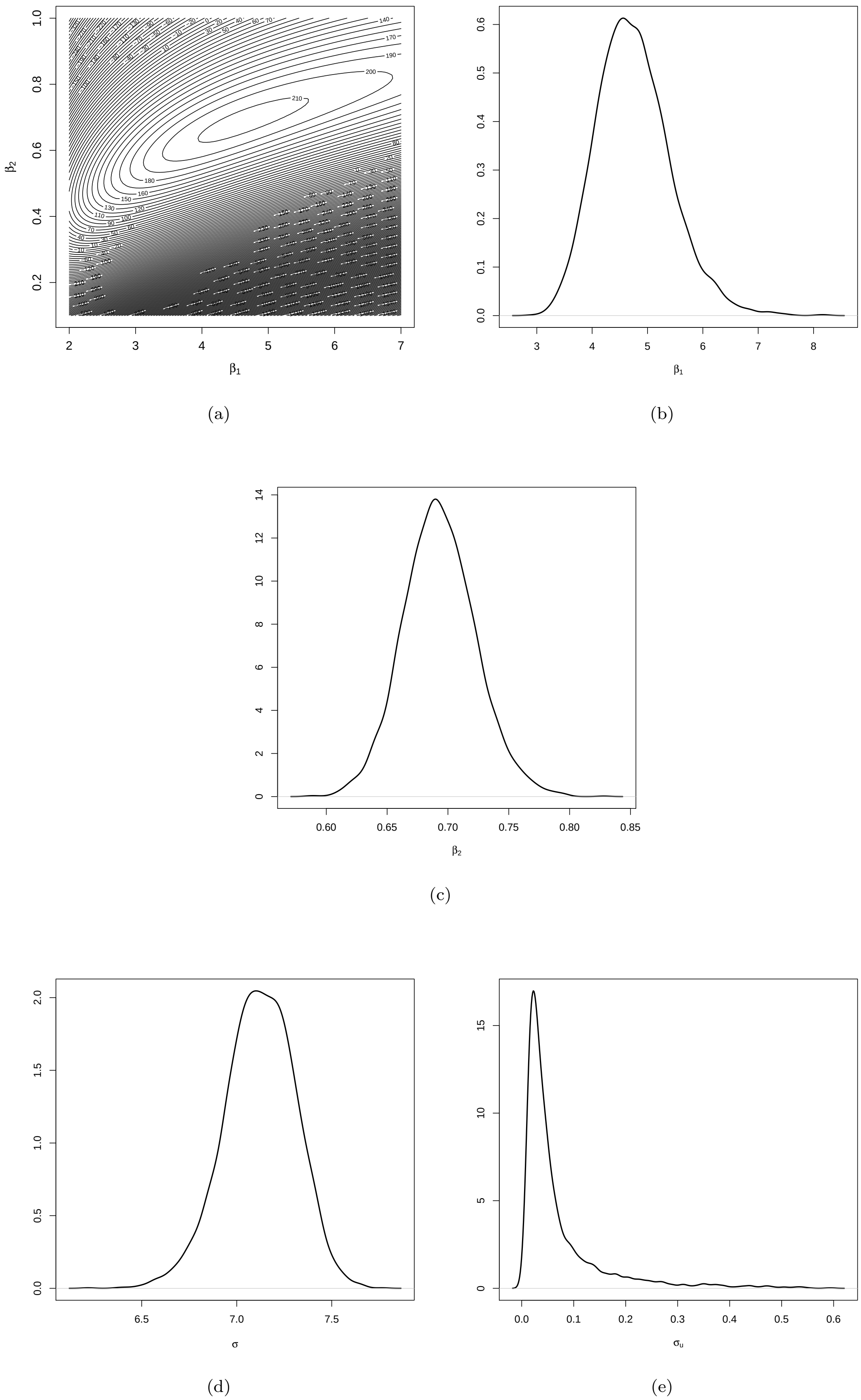


Figure 7. (a) Contour plot of the log posterior density of β_1 and β_2; (b) marginal density distribution of β_1; (c) marginal posterior density of β_2; (d) marginal density distribution of σ; (e) marginal density distribution of σ_u.

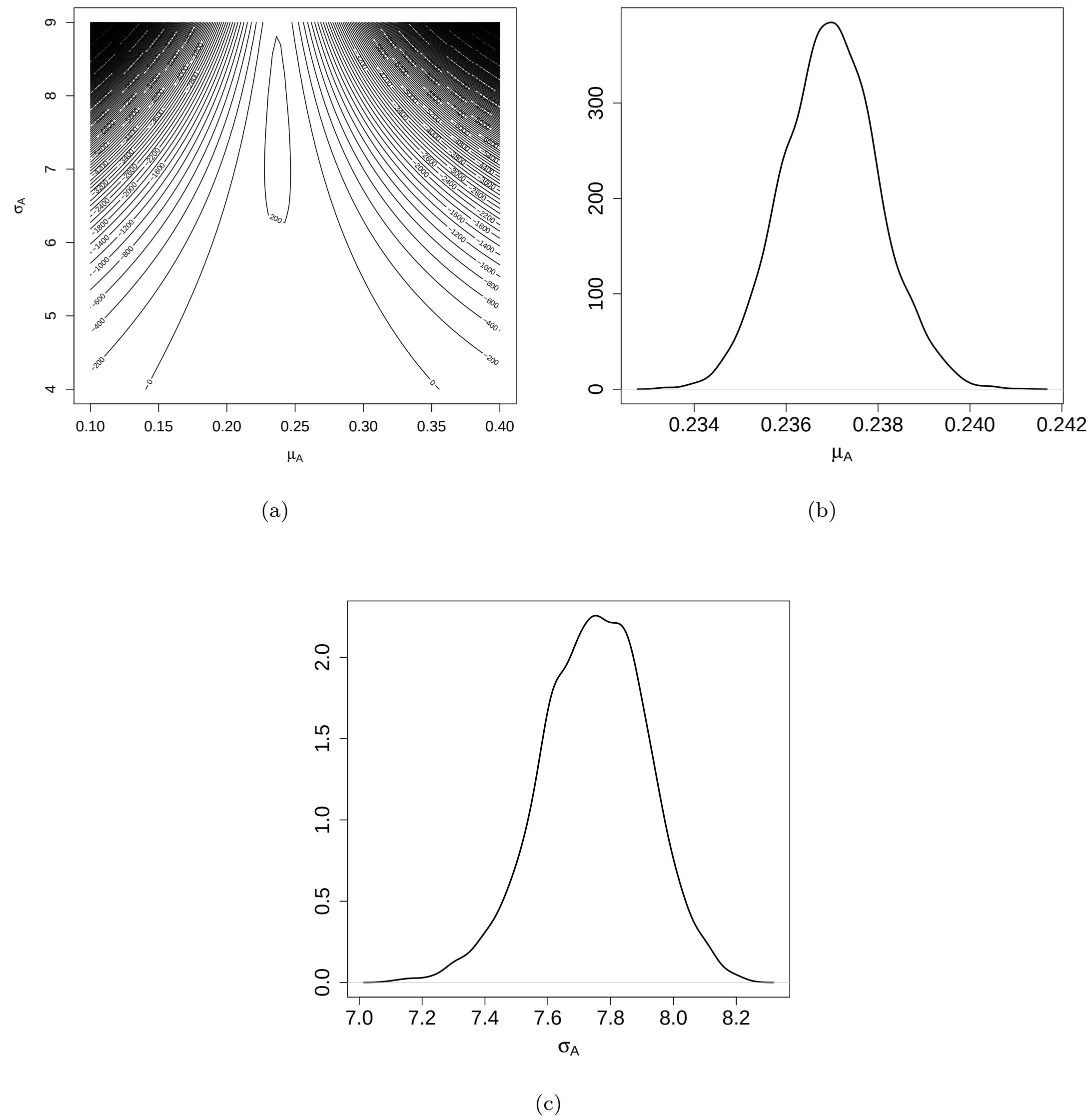
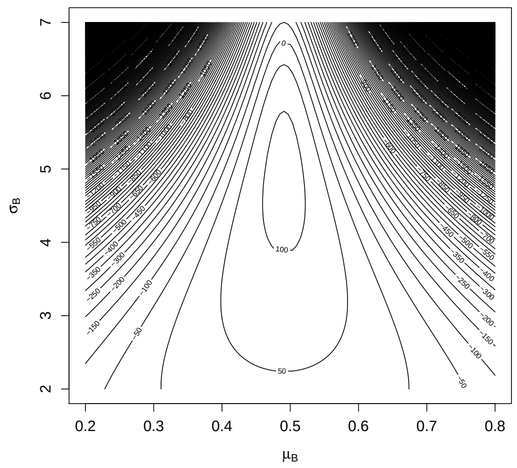
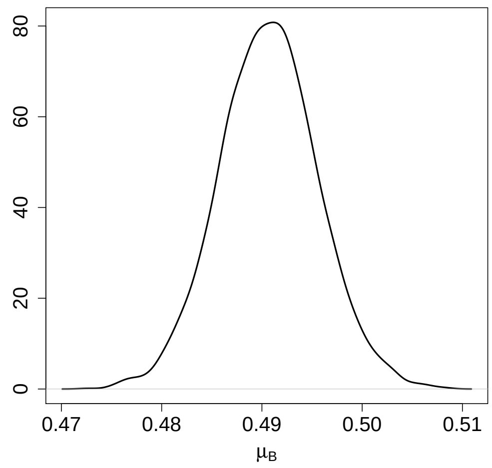


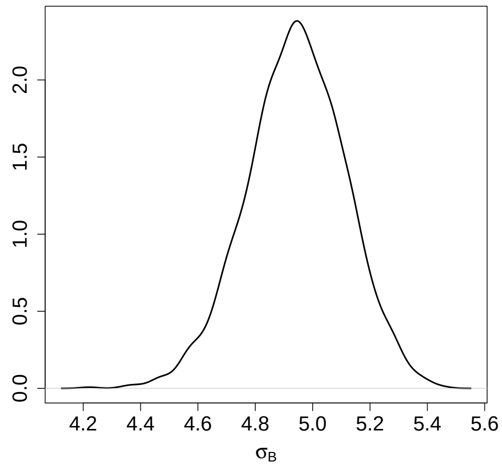
Figure 8. (a) Contour plot of the log posterior density of μ_A and σ_A; (b) marginal posterior density of μ_A; (c) marginal posterior density of σ_A.



(a)



(b)



(c)

Figure 9. (a) Contour plot of the log posterior density of μ_B and σ_B; (b) marginal posterior density of μ_B; (c) marginal posterior density of σ_B.

Table 1. Prior sensitivity analysis: posterior summaries for β_1, β_2, σ, σ_u, u_1, u_2 and u_3 under the for the randomly truncated nonlinear beta mixed-effects model with Gardner's SWCC.

Parameter $\sim$ Prior	True value	Posterior mean	Posterior median	Std. dev.	95% Credible intervals Interquantile	HPD	Acceptance rate
$\beta_1 \sim IG\left(\tau^{-1}, \tau^{-1}\right)$	2.25	1.8859	1.8634	0.2753	(1.4361, 2.4827)	(1.3569, 2.3817)	0.33
$\beta_2 \sim IG\left(\tau^{-1}, \tau^{-1}\right)$	0.5	0.5073	0.5071	0.0242	(0.4591, 0.5574)	(0.4577, 0.5525)	0.63
$\sigma \sim N\left(0, \tau\right)$	6	6.1706	6.1699	0.1691	(5.8357, 6.4989)	(5.8004, 6.4543)	0.71
$\sigma_u \sim IG\left(\tau^{-1}, \tau^{-1}\right)$	0.1	0.0639	0.0498	0.0509	(0.0201, 0.2302)	(0.0156, 0.1681)	0.79
$u_1 \sim N\left(0, \sigma_u\right)$	0.0493	0.0271	0.0277	0.0101	(0.0078, 0.0457)	(0.0071, 0.0447)	0.19
$u_2 \sim N\left(0, \sigma_u\right)$	-0.0226	-0.0320	-0.0314	0.0105	(-0.0511, -0.0131)	(-0.0509, -0.0129)	0.19
$u_3 \sim N\left(0, \sigma_u\right)$	-0.0278	-0.0411	-0.0411	0.0102	(-0.0615, -0.0220)	(-0.0614, -0.0220)	0.19
$\beta_1 \sim IG\left(\tau^{-1}, \tau^{-1}\right)$	2.25	1.8899	1.8645	0.2598	(1.4473, 2.4893)	(1.3681, 2.3825)	0.34
$\beta_2 \sim IG\left(\tau^{-1}, \tau^{-1}\right)$	0.5	0.5078	0.5065	0.0248	(0.4584, 0.5630)	(0.4559, 0.5587)	0.63
$\sigma \sim N\left(0, \tau\right)$	6	6.1597	6.1658	0.1804	(5.7675, 6.5248)	(5.7417, 6.4702)	0.71
$\sigma_u^r \sim t_+\left(\nu, \tau\right)$	0.1	0.1151	0.0813	0.1040	(0.0265, 0.4523)	(0.0137, 0.3355)	0.59
$u_1 \sim N\left(0, \sigma_u\right)$	0.0493	0.0276	0.0276	0.0097	(0.0093, 0.0483)	(0.0078, 0.0468)	0.20
$u_2 \sim N\left(0, \sigma_u\right)$	-0.0226	-0.0322	-0.0313	0.0093	(-0.0527, -0.0155)	(-0.0533, -0.0167)	0.20
$u_3 \sim N\left(0, \sigma_u\right)$	-0.0278	-0.0415	-0.0414	0.0088	(-0.0583, -0.0247)	(-0.0578, -0.0242)	0.20
$\beta_1^r \sim N\left(0, \tau\right)$	2.25	1.9181	1.8905	0.2694	(1.4324, 2.5117)	(1.3784, 2.3923)	0.34
$\beta_1^r \sim N\left(0, \tau\right)$	0.5	0.5065	0.5056	0.0239	(0.4642, 0.5525)	(0.4623, 0.5517)	0.62
$\sigma \sim N\left(0, \tau\right)$	6	6.1725	6.1823	0.1863	(5.8175, 6.4837)	(5.8589, 6.5097)	0.70
$\sigma_u^r \sim N\left(0, \tau\right)$	0.1	0.0493	0.0407	0.0379	(0.0195, 0.1200)	(0.0158, 0.1008)	0.71
$u_1 \sim N\left(0, \sigma_u\right)$	0.0493	0.0288	0.0289	0.0099	(0.0097, 0.0485)	(0.0081, 0.0459)	0.19
$u_2 \sim N\left(0, \sigma_u\right)$	-0.0226	-0.0303	-0.0301	0.0100	(-0.0503, -0.0117)	(-0.0489, -0.0105)	0.19
$u_3 \sim N\left(0, \sigma_u\right)$	-0.0278	-0.0397	-0.0390	0.0095	(-0.0593, -0.0235)	(-0.0593, -0.0235)	0.19

Table 2. Prior sensitivity analysis: posterior summaries for μ_A and σ_A under the for the randomly truncated nonlinear beta mixed-effects model with Gardner's SWCC.

Parameter $\sim$ Priori	True value	Posterior mean	Posterior median	Std. dev.	95% Credible intervals Interquantile	HPD	Acceptance rate
$\mu_A \sim Beta\,(1/2, 1/2)$	0.25	0.2470	0.2467	0.0037	(0.2398, 0.2543)	(0.2393, 0.2536)	0.56
$\sigma_A \sim N\,(0, \tau)$	5.2	5.1554	5.1573	0.1716	(4.8086, 5.4632)	(4.8054, 5.4629)	0.56
$\mu_A^r \sim N\,(0, \tau)$	0.25	0.2474	0.2472	0.0039	(0.2405, 0.2552)	(0.2405, 0.2552)	0.56
$\sigma_A \sim N\,(0, \tau)$	5.2	5.1773	5.1888	0.1747	(4.8332, 5.4978)	(4.8337, 5.5000)	0.56

Table 3. Prior sensitivity analysis: posterior summaries for μ_B and σ_B under the for the randomly truncated nonlinear beta mixed-effects model with Gardner's SWCC.

Parameter $\sim$ Priori	Real value	Posterior mean	Posterior median	Std. dev.	95% Credible intervals Interquantile	HPD	Acceptance rate
$\mu_B \sim Beta\,(1/2, 1/2)$	0.6	0.5985	0.5985	0.0055	(0.5878, 0.6100)	(0.5875, 0.6093)	0.56
$\sigma_B \sim N\,(0, \tau)$	4.6	4.7125	4.7314	0.1700	(4.3781, 5.0387)	(4.3504, 5.0014)	0.56
$\mu_B^r \sim N\,(0, \tau)$	0.6	0.5990	0.5993	0.0052	(0.5883, 0.6101)	(0.5899, 0.6111)	0.56
$\sigma_B \sim N\,(0, \tau)$	4.6	4.7258	4.7254	0.1676	(4.3975, 5.0383)	(4.3977, 5.0386)	0.56

Table 4. Simulation results for the randomly truncated beta Gardner mixed-effects model.

n	Parameter	True value	Posterior mean	Bias-m	MSE-m	Posterior median	Bias-md	MSE-md	Coverage probability Interquantile	HPD	Acceptance rate
	β_1	2.25	2.8249	5.75E-01	1.42E+00	2.3715	5.75E-01	7.10E-01	0.96	0.98	0.40
	β_2	0.50	0.5508	5.08E-02	6.57E-03	0.5335	5.08E-02	3.90E-03	0.90	0.93	0.65
	σ	6.00	5.8761	-1.24E-01	1.26E-01	5.9152	-1.24E-01	1.13E-01	0.96	0.97	0.71
	σ_u	0.10	0.2427	1.43E-01	4.58E-02	0.1318	1.43E-01	1.79E-02	1.00	1.00	0.69
	u_1	-	0.0149	2.83E-03	3.13E-03	0.0124	2.83E-03	3.08E-03	0.99	0.98	0.25
24	u_2	-	0.0084	1.74E-03	2.73E-03	0.0032	1.74E-03	2.68E-03	0.98	0.96	0.25
	u_3	-	0.0045	1.14E-03	3.01E-03	0.0040	1.14E-03	2.96E-03	0.97	0.95	0.25
	μ_A	0.25	0.2493	-7.38E-04	4.44E-05	0.2490	-7.38E-04	4.48E-05	0.95	0.96	0.56
	σ_A	5.20	5.2022	2.18E-03	8.90E-02	5.1942	2.18E-03	8.88E-02	0.95	0.95	0.56
	μ_B	0.60	0.6002	2.44E-04	9.99E-05	0.6002	2.44E-04	9.97E-05	0.96	0.95	0.56
	σ_B	4.60	4.5827	-1.73E-02	1.23E-01	4.5969	-1.73E-02	1.23E-01	0.90	0.90	0.56
	β_1	2.25	2.3639	1.14E-01	2.35E-01	2.2457	1.14E-01	2.08E-01	0.95	0.95	0.32
	β_2	0.50	0.5093	9.30E-03	8.36E-04	0.5069	9.30E-03	7.79E-04	0.97	0.96	0.61
	σ	6.00	5.9638	-3.62E-02	4.04E-02	5.9533	-3.62E-02	4.14E-02	0.94	0.94	0.70
	σ_u	0.10	0.2017	1.02E-01	1.98E-02	0.1158	1.02E-01	2.39E-03	1.00	1.00	0.65
	u_1	-	-0.0027	-4.00E-04	2.33E-03	-0.0035	-4.00E-04	2.32E-03	0.95	0.94	0.18
72	u_2	-	0.0051	-6.03E-04	2.28E-03	0.0033	-6.03E-04	2.27E-03	0.95	0.95	0.18
	u_3	-	0.0062	1.04E-04	2.21E-03	0.0057	1.04E-04	2.20E-03	0.93	0.93	0.18
	μ_A	0.25	0.2498	-1.91E-04	1.44E-05	0.2499	-1.91E-04	1.42E-05	0.93	0.93	0.56
	σ_A	5.20	5.1965	-3.49E-03	3.03E-02	5.1822	-3.49E-03	3.07E-02	0.94	0.93	0.56
	μ_B	0.60	0.5998	-1.71E-04	3.58E-05	0.5998	-1.71E-04	3.59E-05	0.94	0.93	0.56
	σ_B	4.60	4.6050	4.97E-03	3.15E-02	4.6148	4.97E-03	3.15E-02	0.93	0.93	0.56
	β_1	2.25	2.2634	1.34E-02	1.75E-02	2.2605	1.34E-02	1.73E-02	0.92	0.92	0.29
	β_2	0.50	0.5012	1.18E-03	6.70E-05	0.5005	1.18E-03	6.61E-05	0.96	0.96	0.59
	σ	6.00	5.9934	-6.65E-03	2.71E-03	5.9912	-6.65E-03	2.73E-03	0.97	0.97	0.70
	σ_u	0.10	0.2015	1.02E-01	4.54E-02	0.1086	1.02E-01	3.79E-02	1.00	1.00	0.64
	u_1	-	0.0013	-7.80E-06	1.86E-03	-0.0021	-7.80E-06	1.85E-03	0.90	0.90	0.14
720	u_2	-	0.0165	2.36E-04	2.14E-03	0.0226	2.36E-04	2.14E-03	0.92	0.92	0.14
	u_3	-	0.0125	-2.42E-04	1.91E-03	0.0164	-2.42E-04	1.91E-03	0.92	0.91	0.14
	μ_A	0.25	0.2501	1.20E-04	1.46E-06	0.2501	1.20E-04	1.45E-06	0.94	0.95	0.55
	σ_A	5.20	5.1991	-9.14E-04	2.70E-03	5.2024	-9.14E-04	2.70E-03	0.96	0.96	0.55
	μ_B	0.60	0.6001	1.08E-04	3.47E-06	0.6001	1.08E-04	3.50E-06	0.96	0.97	0.55
	σ_B	4.60	4.5967	-3.27E-03	2.71E-03	4.5947	-3.27E-03	2.72E-03	0.95	0.94	0.55

Table 5. Median value of model selection criteria across simulated datasets.

Model	$\tilde{\rho}$	log-CPO
Truncated beta Gardner mixed-effects* (M_0)	0.53	0.47
Truncated beta van Genuchten mixed-effects (M_1)	225.58	220.07
Proportion of correct selection	0.61	0.61

*simulated model.

Table 6. Posterior summary of the randomly truncated beta Gardner mixed-effects model (M_1) fitted to the dataset.

Parameter	Posterior mean	Posterior median	Std. deviation	95% credible intervals Interquantile	HPD
β_1	4.7550	4.7032	0.6638	(3.6256, 6.2120)	(3.5018, 6.0656)
β_2	0.6932	0.6921	0.0299	(0.6371, 0.7555)	(0.6329, 0.7511)
σ	7.1214	7.1265	0.1874	(6.7276, 7.4606)	(6.7535, 7.4825)
σ_u	0.0706	0.0397	0.0846	(0.0106, 0.3492)	(0.0062, 0.2506)
u_1	0.0070	0.0071	0.0054	(-0.0033, 0.0171)	(-0.0032, 0.0172)
u_2	0.0176	0.0176	0.0056	(0.0066, 0.0288)	(0.0070, 0.0290)
u_3	0.0195	0.0194	0.0043	(0.0111, 0.0275)	(0.0110, 0.0274)
μ_A	0.2370	0.2370	0.0010	(0.2350, 0.2391)	(0.2349, 0.2390)
σ_A	7.7456	7.7514	0.1692	(7.3919, 8.0615)	(7.3973, 8.0641)
μ_B	0.4907	0.4907	0.0049	(0.4810, 0.5004)	(0.4813, 0.5007)
σ_B	4.9422	4.9454	0.1715	(4.5887, 5.2734)	(4.6156, 5.2884)